Applications of Fluorine Magnetic Resonance for Small-Molecule Screening, Ligand Development, and Oxygen Sensing

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Clifford T. Gee

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

ADVISOR:
Dr. William C. K. Pomerantz

AUGUST 2017

Acknowledgements

Many thanks to all of my family, friends, coworkers, and collaborators for the unique roles you have each played in my development as a person and as a researcher.

Professor Will Pomerantz - For being a great mentor and advisor over the past five years. It has been quite the journey from setting up the lab to running urban races around Minneapolis to seeing how much the lab has grown and developed.

Professor Gunda Georg – For providing the opportunity for interdisciplinary training and for giving me the opportunity to work alongside some great researchers.

The Pomerantz Lab – For creating such a positive work environment. Everyone may have their own quirks, but it always made for an interesting and entertaining time and helped provide an enjoyable place to work.

Professor Kevin Huang – For recognizing my potential long before I did. Without the encouragement early on in my undergraduate career to pursue research, I would not be where I am now.

Professor Abby Hodges – For putting the field of chemical biology on my radar. Had it not been for that one class senior year, I might not have even considered working at the interface between chemistry and biology.

Dr. Jason Rosé – For inspiring my younger self to enjoy chemistry. Everything from the Pink Sheet to negative extra credit to decimating the AP Chemistry test has helped lead me to be the scientist I am today.

Twin Cities Chinese Christian Church Family and Friends – For being an endless source of encouragement and opportunity for growth. T4C has become like a second home, and many of its members, like family (or in some cases, literal family). The memories of helping to lead the congregation in music, the relationships developed in small group, and the mentoring opportunities through the youth group will not be easily forgotten.

Family – For support, encouragement, and inspiration to purse my interests, wherever they may lead

God – For peace in the midst of trials, hope in the midst of fear, confidence in the mist of doubt, and purpose in the midst of life.

Abstract

Protein-protein interactions (PPIs) play a vital role in biological processes but are difficult to target therapeutically. However, targeting PPIs is an important challenge because their dysregulation is linked to many various disease states including cancers and neurological disorders. While high throughput screening (HTS) has long been the standard method for drug discovery, fragment-based screening (FBS) has emerged as a promising alternative strategy due to its greater coverage of chemical space with smaller library sizes. Successful cases like Vemurafenib and Venetoclax, continue to bolster FBS efforts. Though many techniques, including X-ray crystallography, surface plasmon resonance, and thermal shift assays, have all been used as screening tools, the central hypothesis of this dissertation is that ^{19}F NMR is a powerful and time efficient FBS tool that is complementary to existing tools and is useful for characterizing proteins and small molecule ligands. Protein-Observed Fluorine NMR Spectroscopy (PrOF NMR) due to its high speed, lack of background signals, environmental sensitivity, is an ideal method to use for both ligand discovery and characterization of ligand-protein interactions. Herein, we describe the application of PrOF NMR to two proteins in particular, the KIX domain of CBP/p300 which is part of a larger transcriptional activation complex, and the first bromodomain of BrdT, an epigenetic "reader" protein that has been validated as a target for male contraception. We demonstrate the use of PrOF NMR as a primary screening tool for KIX, identifying key pharmacophores for KIX binding. We also demonstrate the use of PrOF NMR for characterizing ligand-protein interactions, uncovering a new binding site in KIX, distinct from its two native transcription factor binding sites. Validation of hits from other screening campaigns can also be followed via PrOF NMR, and the quantitative information obtained can be used to guide the structure-activity relationship (SAR) process for further ligand development. Beyond ligand discovery in proteins, fluorine magnetic resonance can also be applied as an imaging and oximetry tool. Given the sensitivity of fluorine and its applications in both biophysical and biomedical contexts, fluorine magnetic resonance serves as a new tool for small-molecule screening, ligand development, and oxygen sensing.

Table of Contents

List of Tables

List of Figures

Preface

Chapter 1 of this dissertation provides an overarching background to protein-protein interactions, their significance, and how one can approach small-molecule discovery to specifically modulate these interactions. The KIX domain of CBP/p300 was selected as a model system due to its many native transcription factor binding partners. To study this protein, many biophysical tools, including circular dichroism and fluorescence polarization, were used. Fundamentals of NMR techniques and small-molecule screening are also discussed to set up the subsequent chapters.

In Chapter 2 of this dissertation, the first demonstration of a fragment screen against a protein target using our newly developed technique, protein-observed fluorine NMR spectroscopy (PrOF NMR) is presented. We incorporated 3-fluorotyrosine (3FY) into KIX, our model, system, and screened 508 fragment compounds to assess the feasibility of PrOF NMR as a screening tool. Hits are identified based on changes in chemical shift, and affinities are quantified by plotting the change in chemical shift as a function of ligand concentration. From this screen, we identified aromatic carboxylic acids as a minimum pharmacophore for binding to KIX. We were also able to assess the druggability of the different binding sites based on comparative hit rates and ascertained that the MLL site was more druggable than the CREB site. Due to the speed of the experiments, ease of interpretation, low amount of material consumed, and relatively low cost, we conclude that PrOF NMR is a useful, complementary tool for the screening community.

Chapter 3 is a follow-up to the work presented in Chapter 2 and encompasses two projects that stemmed from the previous work. With the newly discovered hit compounds, we began to synthesize derivatives of the lead compound, 3, 4-dichlorophenyl acetic acid. Because our hypothesis from the initial paper was that this compound was binding to KIX in a non 1:1 manner, we sought to expand the compound while retaining both the dichlorophenyl group on one end and the acidic functionality on the other. Through amide bond coupling with amino acids, we developed a new compound with slightly improved affinity for KIX. Simultaneously, additional efforts were taking place to push the limits of fluorine labeling of proteins by seeking to

incorporate more than one type of fluorinated amino acid into a protein at the same time. These efforts produced KIX labeled with both 3FY and 4-fluorophenylalanine (4FF), providing an additional fluorine probe in the MLL site, the site occupied by the hits from Chapter 2. In putting these two projects together, we discovered a new binding site in KIX, distinct from the two used by native transcription factors. Through a combination of PrOF NMR, ^{1}H-^{15}N Heteronuclear Single Quantum Correlation (HSQC) NMR and computational methods (FTMap, Glide Docking), we propose the existence of a previously unreported ligand binding site.

The protocol in Chapter 3 fully explains the entire process for establishing PrOF NMR in a new context, from the first steps of transforming plasmids into appropriate cells to fragment screening preparation. Using commercially available fluorinated amino acids or fluorinated amino acid precursors, we demonstrate the applicability of PrOF NMR with three different proteins, the KIX domain of CBP/p300, and the first bromodomains of Brd4 and BrdT. PrOF NMR is able to provide both quantitative information, through evaluation of binding affinity, and structural information, based on which resonances are perturbed in the presence of a ligand. This method provides a new tool that is complementary to existing screening tools. While experiments can already be run on the order of a few minutes, we also sought to improve the throughput with the addition of paramagnetic metal additives which shorten relaxation times. Shorter relaxation times allow for shorter experiment times. We identified an optimal metal-chelate for application to PrOF NMR which is able to shorten experiment times by approximately 60%, making PrOF NMR even more efficient and accessible to a wider range of researchers.

In a collaborative effort, PrOF NMR was extended to a new protein target, BrdT, a bromodomain that is a validated target for male contraception and is discussed in Chapter 4. There is currently a market need for non-hormone based male contraceptive agents, and the inhibition of the epigenetic “reader” protein BrdT has been shown to exhibit the desired effects. BrdT knockout mice are viable but sterile, while mice that have been dosed with a bromodomain inhibitor regain spermatogenic function upon removal of treatment. Ligand development for this project took place through multiple

avenues including ligand deconstruction, fragment screening, and virtual screening. The goal in all three cases was to develop a ligand selective for BrdT over the other bromodomains in the same family. By deconstructing a kinase inhibitor that was previously reported to inhibit BrdT and assessing the structure-activity relationship (SAR), we sought to reconstruct a compound that would be selective for BrdT over Brd4. The SAR indicated the possibility of generating selectivity through an N-alkyl substitution pattern. However, the final compound failed to provide the desired selectivity. Fragment screening efforts identified several pharmacophores for binding to the BET bromodomains. The lead compound featured an indanone, a pharmacophore that had not yet been identified. Several analogs of this lead compound were synthesized, and SAR was guided through the use of PrOF NMR by providing affinities, ligand efficiencies, and lipophilic ligand efficiencies. A virtual screening campaign yielded several hits which were then evaluated by PrOF NMR. The lead compound identified by PrOF NMR was then further developed. While final compounds that were developed were not selective for BrdT, this project demonstrated the applicability of virtual screening in conjunction with biophysical assays.

A collaborative project utilizing my expertise in fluorine NMR is described in Chapter 6. Ultraporous mesostructured silica nanoparticles were synthesized and loaded with perfluorocarbons. Fluorine is extremely sensitive to its chemical environment. In particular, dissolved oxygen will significantly modulate the relaxation times of the fluorine nuclei. Fluorine NMR was used to both quantify the amount of perfluorocarbon that had been loaded into the particles and to monitor changes in relaxation as a function of oxygen concentration. Our idea was to take advantage of this property and use these loaded nanoparticles, termed PERFUMNS, as oximetry sensors with applications in the biomedical field. In collaboration with the Center for Magnetic Resonance Research (CMRR) we were able to obtain ^{19}F MRI images of particles in solution demonstrating clinically-relevant concentrations of fluorine for imaging experiments.

The final chapter, Chapter 7, describes a new method for using standard colorimetric assays to measure protein concentrations. Smartphone colorimetry uses image analysis of the Red, Green, Blue (RGB) values in the place of absorbance in a UV-

Vis spectrophotometer. We demonstrate the applicability of smartphone colorimetry in two different protein colorimetric assays, the Bradford assay and the biuret assay, and provide instructions for both students and instructors to be able to utilize this experiment as part of a laboratory course, meeting some Next Generation Science Standards as well.

In summary, the findings presented in this dissertation describe my contributions to the ligand discovery field through the development of protein-observed fluorine NMR, to the biomedical field through the development of a new oximetry measurement tool, and to the field of chemical education the development of a new application of smartphone colorimetry to protein-based assays.

Citations from previously published work in this dissertation are as follows:

Chapter 2:

"Fragment Screening and Druggability Assessment for the CBP/p300 KIX Domain through Protein-Observed 19F NMR," C. T. Gee, E. J. Koleski, W. C. Pomerantz, *Angewandte Chemie International Editio* **2015**, *54* (12), 3734-3739.

Chapter 4:

"Protein-observed ^{19}F-NMR for fragment screening, affinity quantification and druggability assessment," C. T. Gee, K. E. Arntson, A. K. Urick, N. K. Mishra, L. M. L. Hawk, A. J. Wisniewski, W. C. K. Pomerantz, *Nature Protocols* **2016**, 1414-1427.

*"*Paramagnetic relaxation enhancement for protein-observed ^{19}F NMR as an enabling approach for efficient fragment screening," L. M. L. Hawk, C. T. Gee, A. K. Urick, H. Hu, W. C. K. Pomerantz, *RSC Advances* **2016**, 95715-95721.

Chapter 5:

"BET Bromodomain Inhibitors with One-Step Synthesis Discovered from Virtual Screen," A.M. Ayoub, L.M.L. Hawk, R.J. Herzig, J. Jiang, A.J. Wisniewski, C.T. Gee, P. Zhao, J. Zhu, N. Berndt, N.K. Offei-Addo, T.G. Scott, J. Qi, J.E. Bradner, T.R. Ward, E. Schonbrunn, G.I. Georg, and W.C.K. Pomerantz. *Journal of Medicinal Chemistry*, **2017**, *60* (12), 4805-4817.

Chapter 6:

"Oxygen Sensing with Perfluorocarbon-Loaded Ultraporous Mesostructured Silica Nanoparticles," A.L. Lee, C.T. Gee, B.P. Weegman, S.A. Einstein, A.R. Juelfs, H.L. Ring, K. R. Hurley, S.M. Egger, G. Swindlehurst, M.G. Garwood, **W.C.K. Pomerantz**, and C.L. Haynes *ACS Nano*, **2017,** *11* (6), 5623-5632

Chapter 7:

"Quantifying Protein Concentrations Using Smartphone Colorimetry: A New Method for an Established Test," C. T. Gee, E. Kehoe, W. C. K. Pomerantz, R. L. Penn, J. Chem. Educ., **2017**, *94* (7), 941-945.

Chapter 1. Introduction

1.1 **Protein-Protein Interactions**

Protein-protein interactions, long classified as "undruggable" are now being targeted for their roles in various diseases.[1] Under normal, healthy conditions, the putative 650,000 PPIs in the human body are vital for biological function, including the regulation of cell cycle and gene transcription.[2] However, PPIs have also been implicated as drug targets in various diseases including cancer when they are dysregulated. Consequently, the ability to selectively and effectively target these interfaces is highly sought after.[1, 3-4]

In recent years there have been some successes in targeting PPIs with small molecules. Notable examples of PPIs that have been successfully targeted with small molecules include MDM2:p53,[5-7] IL-2:IL-2 Rα,[8-10] and Bcl-2[11-13] (Figure 1.1). These and other successes have been highlighted in recent reviews.[14,15] Despite the rapid progress that has been made in the past 15 years, less than 0.01% of all PPIs in the human body have targeted inhibitors.[16] In contrast, there are about 5,000 different human diseases of which the molecular cause is known.[17] In light of these staggering statistics, clearly more work must be done to optimize the existing small molecules and discover new molecules for targeting these unreached PPIs.

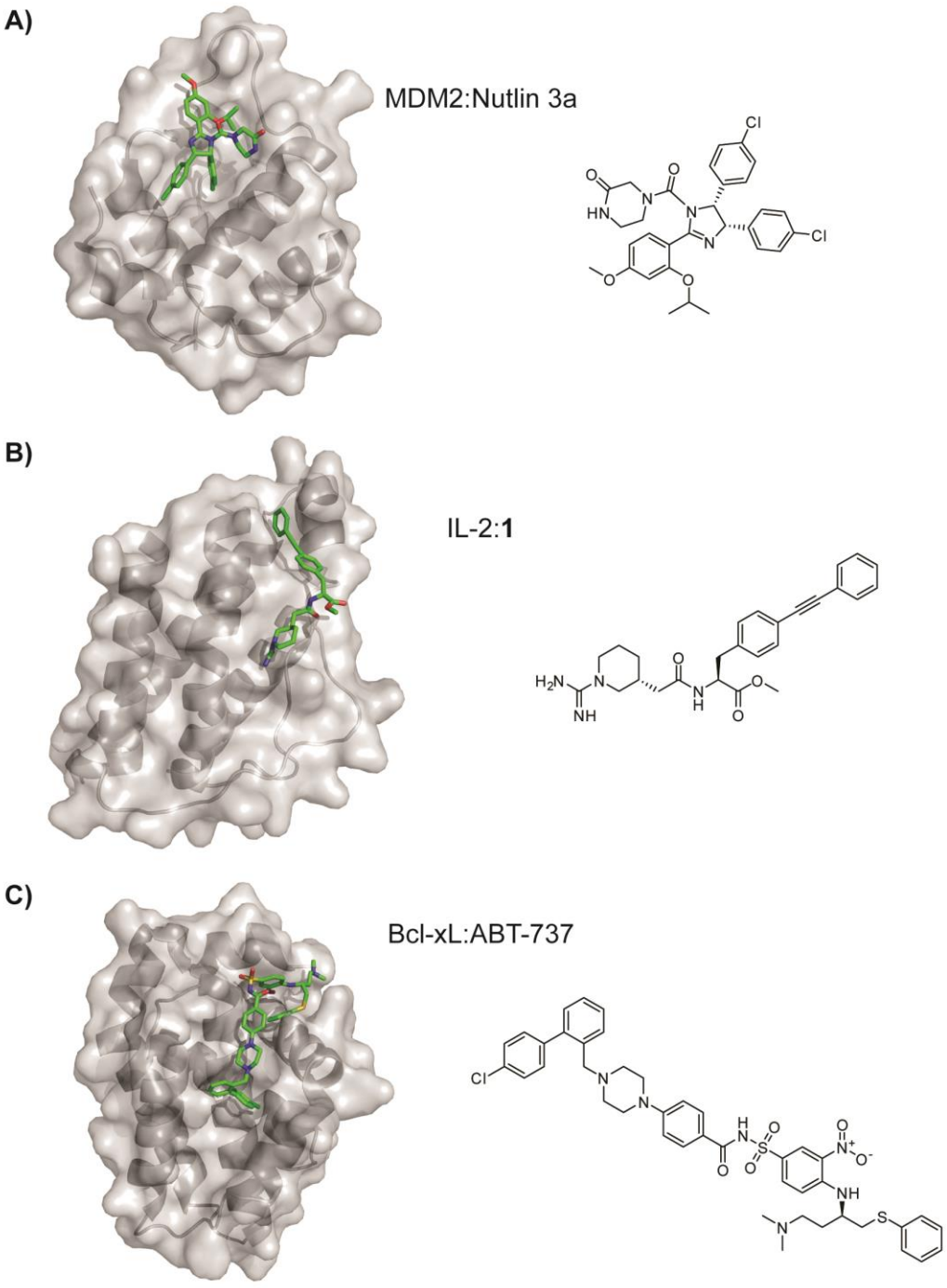


Figure 1.1 Notable examples of small-molecule:PPI inhibitors
A) MDM2:Nutlin 3a (PDB ID: 4HG7) B) IL-2:**1** (PDB ID: 1M48) C) Bcl-xL:ABT-737 (PDB ID: 2YXJ)

1.1.1 Protein Hot Spots

The interface between two proteins is often characterized as both large (1500-3000 $Å^2$ in total) and flat, lacking deep, hydrophobic pockets that could be considered as druggable sites.[18] However, Clackson and Wells, in their seminal paper describing the interface between human growth hormone (hGH) and its receptor, hGHbp, were able to demonstrate through alanine mutagenesis that some amino acids are more important for binding than others. These protein "hot spots" are small regions within the interface, consisting of just a few residues which are essential for protein binding as demonstrated by the large change in binding free energy when mutated.[19] While targeting the entire interface between two proteins may not be feasible, targeting a hot spot to modulate binding interactions is a reasonable task.[20] Notably, in some cases protein hot spots are clustered together while in other cases hot spots are more distributed, which can make it more difficult to target these interactions.

Further investigation into protein hot spots have been carried out to determine which amino acids are overrepresented in hotspots. It was determined that the aromatic amino acids are enriched at protein-protein interfaces as evidenced by their overrepresentation in protein hotspots (Figure 1.2).[21-25] If these hot spot residues could be labeled without perturbing the structure and function of the protein, one would have useful probes in position to monitor protein binding.

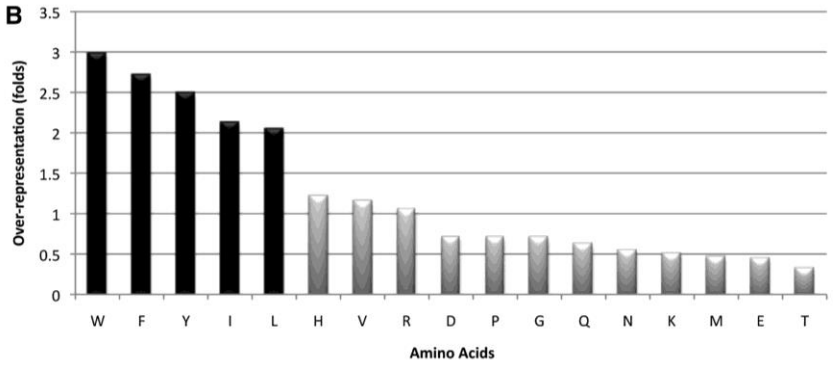


Figure 1.2 Overrepresentation of aromatic amino acids at protein binding interfaces.
Adapted from London et al.[21] with permission from Elsevier.

1.1.2 Transcription Factors and Activators

Transcription factors are of particular interest to cancer research due to their direct role in gene transcription.[26] Transcription factors generally possess at least two distinct domains: a DNA binding domain (DBD) and a transcriptional activation domain (TAD).[27] The DBD binds to specific recognition sequences in the DNA in order to activate transcriptions. This process requires the coordination of many proteins, termed the enhanceosome. One of the key components in the enhanceosome is a coactivator, which binds to its corresponding transcription factor at the TAD to form a bridge with the RNA polymerase II to promote transcription.[28]

Transcription factors, due to their critical role in the regulation and transcription of genes, are therefore valuable therapeutic targets. In the case of cancer, genes are dysregulated to promote continued cell growth and prevent cell death. By targeting the specific transcription factors involved in the PPIs responsible for regulating these genes, one could selectively shut down cancer cell growth.[29] However, because transcription factors are involved in several different PPIs, effective targeted therapeutics would need to inhibit specific PPIs while not perturbing other beneficial functions of the proteins, allowing for normal function of the proteins.

1.1.3 CREB-CBP: a model transcription factor-coactivator interaction

The cyclic adenosine monophosphate (cAMP) response element (CRE) binding protein (CREB) is a biologically important protein due to its implications in neuronal plasticity and long term memory formation as well as its role in hematopoiesis, immune function, cell cycle regulation, gene transcription, and consequently, cancer.[30-34] This 43 kDa protein is a transcription factor that is conserved across multiple species from insects to humans, indicating its significance for biological function.[35]

CREB was first discovered as a transcription factor related to long term memory formation and has now been established to have than 100 target genes controlling a variety of functions including neurotransmission, cell growth, neuronal structure, cellular metabolism, and signal transduction.[34] Each of these genes possesses the same octanucleotide cAMP response element (CRE) sequence, TGANNTCA where N represents any of the four nucleotides.[29, 35] CREB binds as a homodimer to the CRE

sequence and its TAD is subsequently activated by phosphorylation which is performed by various kinases, including protein kinase A and calmodulin-dependent kinase. Upon phosphorylation at Ser133, CREB is able to recruit its coactivator, CREB binding protein (CBP).[36] This binding event between the two proteins allows for the recruitment of RNA polymerase II (RNApolII) to activate transcription of target genes [29-31, 35, 37] (Figure 1.3).

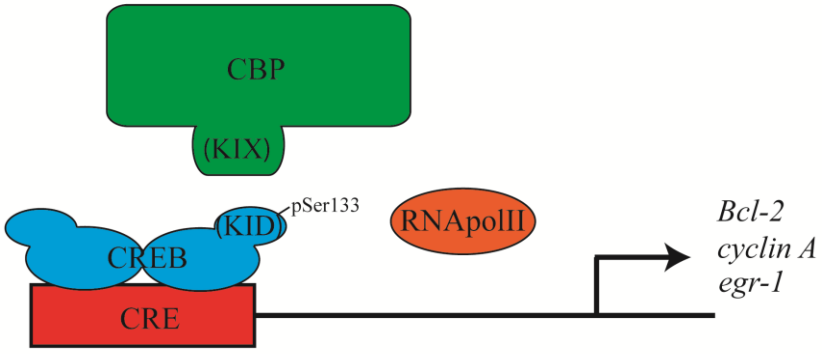


Figure 1.3 Cartoon schematic of CREB:CBP Transcriptional Complex.
CREB binds as a homodimer to the CRE sequence. Interaction between the KID domain of CREB and the KIX domain of CBP is activated by phosphorylation of KID at Ser133, allowing the recruitment of RNApolII for transcription of target genes such as *Bcl-2*, *cyclin A*, and *egr-1*.

CREB has been implicated in the proliferation of AML in which CREB's dysregulation leads to improper transcription of several target genes including the antiapoptotic gene *Bcl-2*, the cell cycle regulating gene *cyclin A*, and the differentiation gene early growth factor-1 (*egr-1)*. [29, 35] With *Bcl*-2 upregulated, apoptosis is inhibited, allowing cells to survive.[11, 13] Likewise the upregulation of cyclin A allows cells to bypass the natural death stage and continue proliferating. The dysregulation of e*gr-1*, causes cells to differentiate improperly, meaning that instead of cells becoming blood cells or neuronal cells, they become cancer cells.[38] Because CREB holds an integral role in the proliferation of leukemia cells, it is an attractive, albeit a difficult, therapeutic target. [29, 35]

1.1.4 KIX – Structure and Function

The CREB-CBP interaction takes place through specific binding domains. CREB features a kinase inducible domain (KID) which is phosphorylated at Ser133, hereafter referred to as pKID. The 256 kDa protein CBP features a 12 kDa KID interacting (KIX) domain, which can be expressed as a minimal protein construct modeling interactions of CREB at the KIX domain. KIX also binds several other transcriptional activators in addition to CREB, including BRCA1, c-Jun, c-Myb, mixed lineage leukemia (MLL) and human papillomavirus E2.[36, 39-40]

The solution structure of KIX consists of three α helices (α1, α2, α3) and two 3_{10} helices. Two α helices, α1 and α3, are coplanar and interact with two of the α helices from pKID, αA and αB (Fig) which form a 90° angle around α3. In contrast, the MLL binding site lies between α2 and α3 away from α1 while c-myb has been found to bind in a similar site as KID (Figure 1.4). [40]

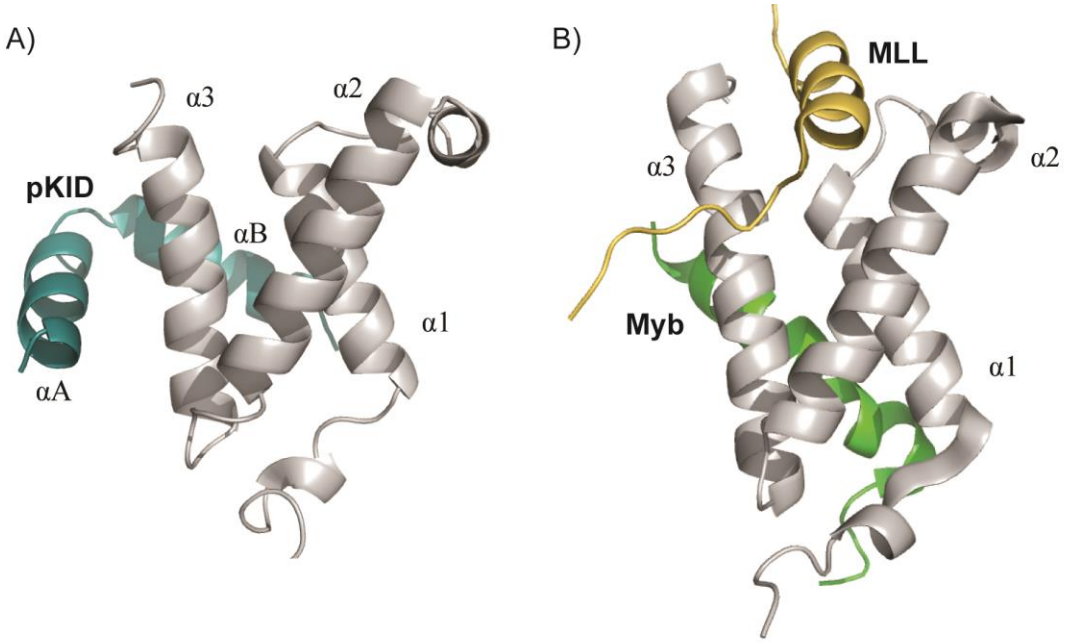


Figure 1.4 Solution structure of KIX with binding partners.
(A) A ribbon diagram of KIX bound to pKID (PDB ID: 1KDX). (B) A ribbon diagram of KIX bound to both MLL and Myb (PDB ID:2AGH) (Structures were generated in PyMOL)

1.1.5 Known Small Molecule and Peptide Ligands for KIX

PPIs involving KIX, unlike most PPIs, do have some known, well-characterized small-molecule and peptide ligands. One of the most prominent small-molecule inhibitors of the KIX-KID interaction is naphthol AS-E phosphate, also known as KG-501,[36] which selectively inhibits the KIX-KID interaction with an initially reported K_i of 90 μM. Other known KIX ligands are listed in Table 1.1.

Table 1.1 Known small-molecule and peptide KIX Ligands

Compound	Inhibition	Affinity	Method/Technique	Citation
Naphthol AS-E Phosphate (KG-501)	KID	90 μM † 115 ± 15 μM *	FBS/SAR by NMR	Best et al.[36] Pomerantz et al.[41]
Sekikaic Acid	MLL KID	34 μM ‡ 64 μM ‡	HTS/FP	Majmudar et al.[42]
Lobaric Acid	MLL KID	17 μM ‡ 25 μM ‡		Majmudar et al.[42]
1G7		----------	FBS/PrOF NMR	Pomerantz et al.[41]
1-10		220 ± 50 μM *	FBS/Tethering (MS)	Pomerantz et al.[41]
KBP 2.20		16.1 μM*	Phage Display	Frangioni et al.[43]

Affinity reported as K_d, K_i, or IC_{50}. (†K_i, *K_d, ‡IC_{50}). Lobaric acid was derived as a conformationally constrained version of Sekikaic acid based on computational modeling which predicted improved binding

While there has been significant investigation into identifying KIX-KID inhibitors [44-46] as well as inhibitors for MLL[47-48] and Jun[48], there is still a need for identifying more high-affinity molecules that demonstrate significant efficacy in selectively targeting the interactions between KIX and its binding partners.

1.1.6 Current Methods to Observe and Characterize Protein-Protein Interactions

Several different yet complementary characterization methods are available for characterizing and obtaining binding information regarding both PPIs and protein-ligand interactions. X-ray crystallography provides angstrom level image resolution of a protein, but the drawback is that this snapshot may not accurately represent its behavior in solution. Forster resonance energy transfer (FRET) can provide useful binding information but necessitates the use of bulky fluorophores which may alter natural protein function. Surface plasmon resonance (SPR), while being label-free, requires the

target protein to be immobilized on a surface which could lead to unnatural binding and function. Isothermal titration calorimetry (ITC) uniquely enables the simultaneous determination of binding constants and thermodynamics by measuring the heat release associated with binding but lacks structural information.[49] For these reasons, multiple characterization techniques are used in tandem to provide a fuller understanding of protein binding.

Fluorescence polarization (FP) is a fluorescence-based technique that monitors the change in the light's polarization based on molecular size and tumbling rate in solution. A smaller molecule that is tumbling faster will exhibit low polarization while a large molecule that tumbles slower will exhibit high polarization. A fluorophore-labeled molecule or peptide that binds to a protein can take advantage of this effect as it will exhibit a significant change in polarization when it is bound versus when it is unbound. [50] For this reason, FP is often used to determine binding constants for protein-ligand binding. These constants can either be found by direct binding experiments yielding the K_d or indirect competition experiments, yielding the K_i. In either case, a fluorophore is attached to a tracer molecule that will bind to the target. In a direct binding experiment, the K_d of a tracer is determined by observing the change in polarization as it relates to its concentration of the tracer. At a high concentration of tracer, low polarization will be detected while at low concentrations, there will be high polarization. (Figure 1.5A, B, Equation 1) For an indirect competition experiment, the tracer is used as a way to measure the binding affinity of another ligand which will displace the tracer from binding to the target protein. As the concentration of the molecule of interest increases, it displaces more of the tracer resulting in a decrease in polarization (Figure 1.5C, D, Equation 2).[50]

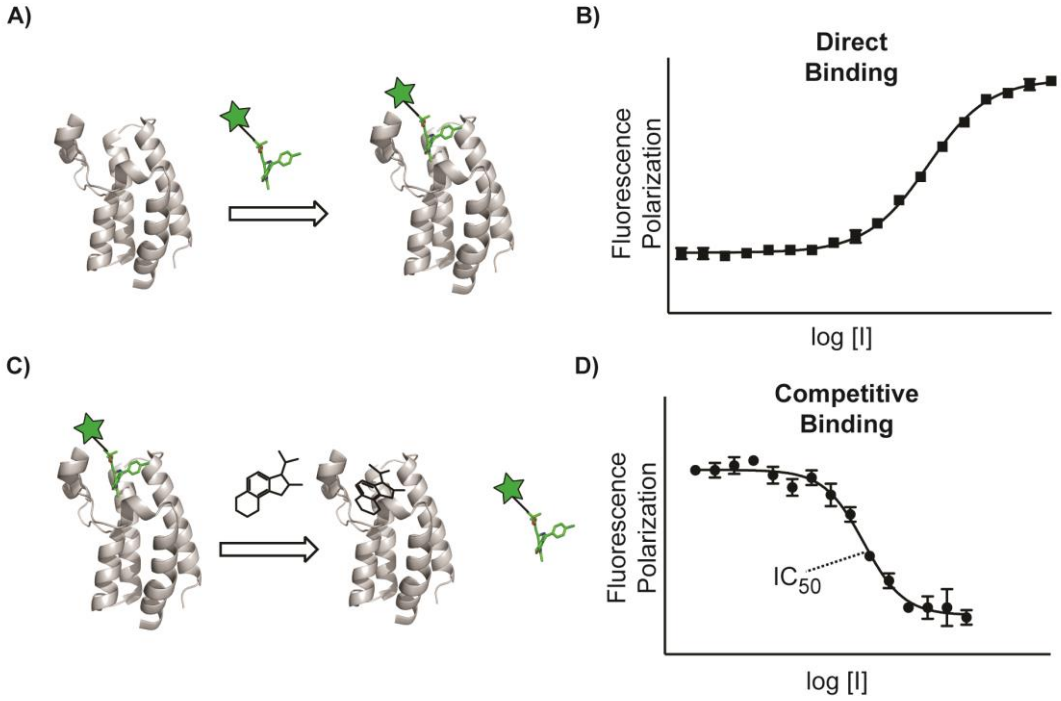


Figure 1.5 General schematic for fluorescence polarization experiments.
A) Schematic of a direct binding experiment in which a fluorescently-labeled ligand is added to a protein, leading to an increase in fluorescence polarization. B) Sample direct binding curve by FP. C) Schematic of a competition experiment in which a fluorescently-labeled ligand is competed away from its protein target by an inhibitor molecule, leading to a decrease in fluorescence polarization. D) Sample FP competition curve with the IC_{50} being noted at the inflection point of the curve.

K_d values in a direct binding experiment are calculated using the equation below:

$$y = c + (b - c) * \frac{\sqrt{(K_d+a+x)^2-4ax}}{2a} \quad (1)$$

In the above equation, y corresponds to the observed anisotropy, c and b correspond to the minimum and maximum observed entropies, respectively, K_d is the dissociation constant, and a and x are the concentrations of fluorescent ligand and protein, respectively. The K_i can be calculated from the IC_{50} values using a variant of the Cheng-Prusoff equation where the K_i is the indirectly determined dissociation constant of the competitor ligand.[51]

$$K_i = \frac{(L_b)(IC_{50})(K_d)}{(L_0)(R_0)+[L_0(R_0+L_b-L_0-K_d)]} \quad (2)$$

Each of the aforementioned techniques has strengths and weaknesses which typically necessitates the use of complementary methods. A further explanation of these and other techniques for small-molecule screening will be discussed in section 1.4. However, the following section will describe how Nuclear Magnetic Resonance may serve as an ideal technique by simultaneously providing structural information, allowing the evaluation of binding affinity, and accurately reflecting solution conditions.

1.2 **Nuclear Magnetic Resonance Spectroscopy**

1.2.1 Protein NMR

Nuclear magnetic resonance (NMR) spectroscopy is an effective tool for analyzing not only small molecules but also large biomolecules like proteins. High field NMR spectrometers have made biomolecular structural studies via NMR not only possible but efficient as well. More powerful magnets provide a higher signal-to-noise-ratio (S/N), allowing for shorter data acquisition or higher resolution data in the same amount of time. The development of cryogenically cooled probes has also provided significant S/N improvement by reducing thermal noise. Cryoprobes increase the S/N three to four-fold. Because S/N is proportional to the square root of the number of scans, a cryoprobe allows data acquisition nine to sixteen times faster than a typical probe. Alternatively, one could have a sample that is significantly less concentrated (mid μM range) and still acquire good quality data in a reasonable amount of time.[52]

Since the advent of the two-dimensional (2D) NMR experiments in 1971, researchers have found numerous uses for application multidimensional NMR in providing structural information.[53] 2D NMR experiments can be classified in one of two categories: correlations between the same types of nuclei (homonuclear) or different types of nuclei (heteronuclear). For proteins, heteronuclear single-quantum correlation (HSQC) or Nuclear Overhauser Effect spectroscopy (NOESY) are often used. Most commonly the HSQC experiments examine the correlation between ^{15}N in the amide backbone and its corresponding ^{1}H or the correlation between ^{1}H and ^{13}C in the sidechains. NOESY determines coupling between two nuclei through space, making it a useful tool for secondary structure determination in proteins.[54] To obtain even more

detailed structural information, three dimensional (3D) and four dimensional (4D) NMR experiments were developed to analyze the correlations among three or four distinct nuclei. This tool revolutionized the protein NMR field, allowing for the acquisition of even more structural information.[55]

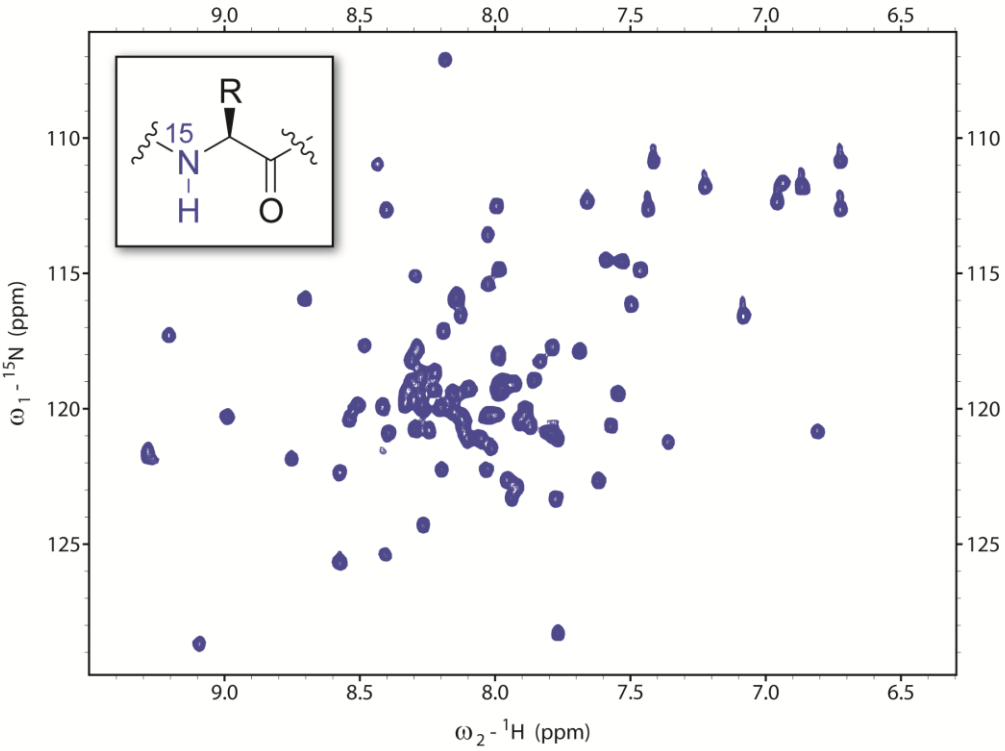


Figure 1.6 Sample ^{1}H-^{15}N HSQC Spectrum of KIX

Traditionally protein observed NMR has studied either the backbone or the side chains. The ^{1}H-^{15}N HSQC experiment is performed by isotopically labeling each of the nitrogen atoms in the amide backbone and observing the correlation to its corresponding hydrogen, providing the overall framework of the protein (Figure 1.6).[56] Two key developments have aided in the more effective application of HSQC experiments. The first is the band-selective optimized flip-angle short-transient (SOFAST) pulse sequence in which the band-selective pulses and short recycle delays allow for faster experiments without loss of S/N.[57]

A powerful innovation was the invention of the transverse-relaxation optimized spectroscopy (TROSY) pulse sequence which makes characterization and structural

determination of larger proteins (> 60 kDa) a reasonable task.[56, 58-59] Under normal conditions, it is difficult to determine the structure of large proteins via NMR due to complex coupling signals that often obscure the important peaks. Due to dipolar coupling and chemical shift anisotropy, relaxation times are also significantly affected as a function of molecular size leading to broad resonances. By eliminating the cross-correlated components with a fast relaxation and retaining the ones with a slow relaxation, TROSY provides improved peak resolution and better quality spectra overall.[56] By coupling TROSY with other NMR experiments like HSQC as well as NOESY and HNCA, large proteins or protein complexes as high as 900 kDa can be analyzed.[58, 60]

Protein NMR can be applied not only for structural determination but also for small molecule discovery. Fesik et al. developed the technique known as structure-activity relationship (SAR) by NMR to identify potential leads for drug discovery.[61] SAR by NMR, which uses the HSQC experiment, was able to identify a trimethoxyphenyl pipecolinic acid derivative and a benzanilide derivative that bound to the FK506 binding protein.[61] Once two ligands that bind nearby sites in the protein, as judged by perturbations in the HSQC NMR spectra, the two fragments can be chemical linked and functionalized, resulting in a synergistic effect on binding affinity in the final compound.[61]

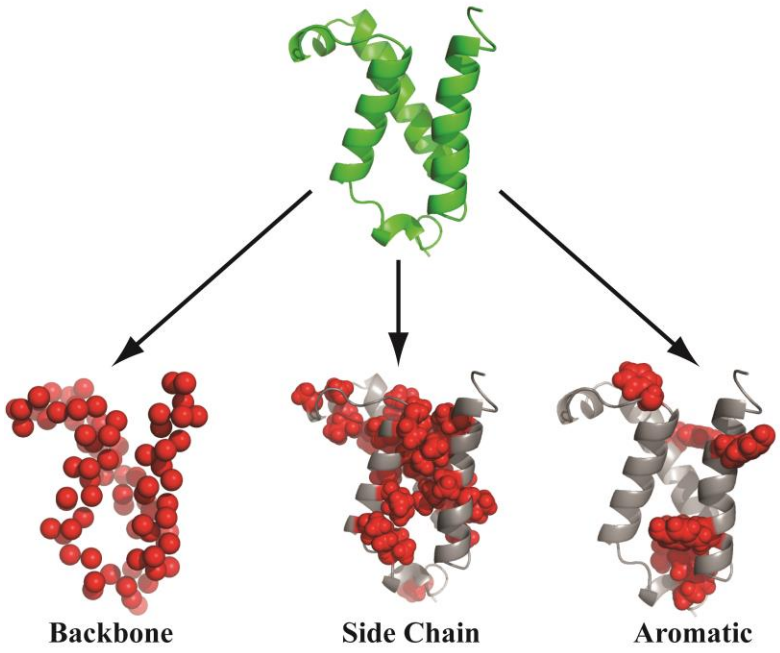


Figure 1.7 Protein labeling for NMR.
(Structures were generated in PyMOL, PDB: 1KDX)

The alternative to isotopically labeling the backbone with ^{15}N is to label the side chains with ^{13}C. Most commonly, the five side chains that are labeled are alanine, valine, leucine, isoleucine, and methionine due to the prominent methyl group in each which allows for methyl-TROSY NMR experiments to be performed.[62-65] Protein characterization through methyl group analysis is advantageous due to their ability to probe structure[66-68] and dynamics.[69] Furthermore, the acquired spectra are well resolved and feature residues that are spread throughout the hydrophobic core of a given protein.[62] Each of the protein labeling techniques (Figure 1.7) has its benefits and provides complementary information for protein structure and function.

1.2.2 Fluorine (^{19}F) NMR

While many advances have been made in NMR, wide application 2D and 3D experiments are still limited by their long experimental times and complicated results. Because these experiments often rely on correlation to ^{13}C or ^{15}N, they are inherently slow due to the need for a high number of scans.[70-71] This temporal challenge has been

diminished by advanced pulse sequences such as SOFAST-HMQC and magnetization transfers allowing for detection of ^{13}C or ^{15}N signals through the proton channel.

In recent years, fluorine (^{19}F) has become more popular due to its powerfully sensitive and versatile nature as an NMR active nucleus.[52] The ^{19}F nucleus is 100% in natural abundance with a gyromagnetic ratio of 25.1815 x 10^7 rad $T^{-1}s^{-1}$, making it 83% as sensitive as ^1H and the second most signal-sensitive NMR active nucleus. It also has a spin quantum number of ½, which means unlike quadrupolar nuclei (spin > ½), it only has two possible spin states which keeps interpretation relatively simple.[71] The Bruker Corporation regards ^{19}F as one of the three most important NMR active nuclei along with ^1H and ^{13}C.[72] Furthermore, because fluorine is not found in natural biomolecules or in physiological conditions, there are no background signals to obfuscate the ^{19}F NMR spectrum. In addition fluorine's chemical shift window is significantly larger than that of either ^1H or ^{13}C. While the chemical shift range for ^1H and ^{13}C is typically 12 and 220 ppm respectively,[54] ^{19}F can span nearly 800 ppm,[73] allowing for widely separated chemical shifts corresponding to fluorine's hyper-responsiveness to slight changes in its chemical environment.[74] Urick et al. reported a 6-20 fold higher responsiveness of fluorine to ligand binding, relative to ^1H.[75] For these reasons, ^{19}F NMR can be used as a powerful characterization tool.

1.2.3 Protein Observed Fluorine NMR (PrOF NMR)

While the term protein observed fluorine NMR (PrOF) NMR was coined recently, proteins have been fluorine-labeled and analyzed previously.[74] Due to fluorine's hyper-responsiveness to changes in its chemical environment, fluorinated amino acids within a protein can serve as effective molecular probes. Even subtle changes in a protein's tertiary structure can result in drastic differences between chemical shifts for the same fluorinated amino acid in different positions within the protein. One such example is the ^{19}F NMR spectrum of the fatty acid binding protein (Figure 1.8), which was labeled with both 4-fluoro-phenylalanine (4FF) and 6-fluoro-tryptophan.[76] Each 4FF residue is distinguished from the other over a 10 ppm window. One key benefit PrOF NMR holds over other protein NMR techniques is the fact that it is a one-dimensional technique, making it inherently simpler to interpret, easier to acquire.[74] While other NMR

techniques provide more structural information, PrOF NMR can serve as a complementary technique to supplement characterization data.

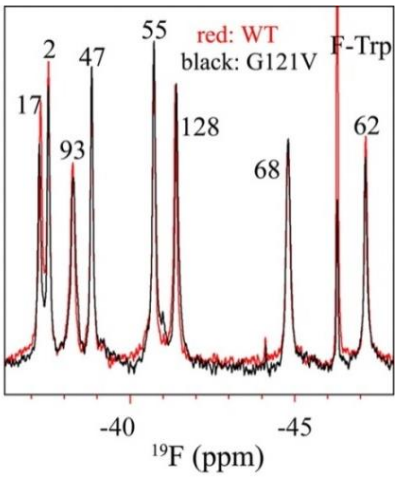


Figure 1.8 ^{19}F NMR spectrum of 4FF/6FW labeled IFABP.
Adapted from Li et al.[76] Copyright 2007 National Academy of Sciences

Because aromatic amino acids are highly enriched at PPI interfaces, they are well suited positionally to serve as NMR probes. By incorporating fluorinated aromatic amino acids, activity at protein-protein interfaces, including protein-protein and protein-small molecule binding, can be monitored. PrOF NMR can also be used to generate unique binding footprints for a protein depending on the binding partner. The 3FY ^{19}F NMR spectra from KIX bound with MLL, pKID, and c-Myb have been reported in the literature in which each spectrum is clearly distinguishable from the other based on the shifting or broadening of resonances (Figure 1.9).[41]

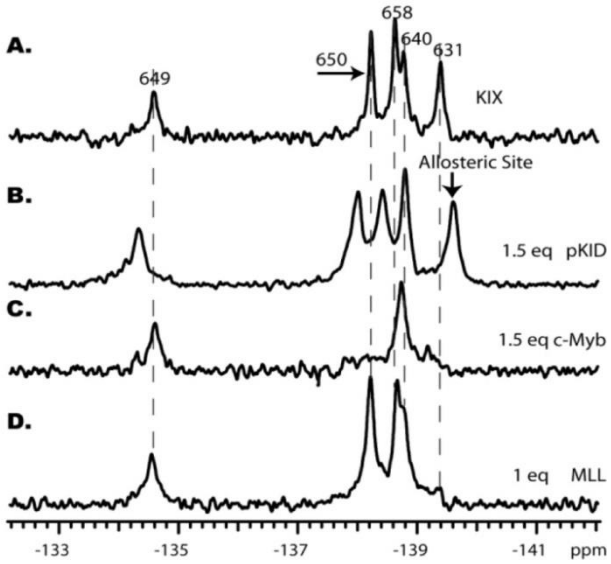


Figure 1.9 ^{19}F NMR of 3FY KIX with different binding partners MLL, c-Myb, and pKID.
Adapted with permission from Pomerantz et al.[41] Copyright 2012 American Chemical Society

The use of PrOF NMR is not limited to a few isolated cases. Rather, it has been used to study a wide variety of proteins including receptor proteins, DNA binding proteins, membrane proteins, enzymes, and even some disordered proteins.[77-78] Some membrane proteins whose structure and function have been characterized by PrOF NMR include the β_2-adrenergic receptor,[79-82] rhodopsin,[83-84] gramicidin A,[85-86] and the M13 phage coat protein.[87-88] For their role in catalyzing various biological processes, enzymes such as purine nucleoside phosphorylase and adenosine deaminase, have been characterized using PrOF NMR. Likewise, disordered proteins, like the Aβ protein aggregate[78] and the SH3 domain of the *Drosophila* signal transduction protein,[89] have been characterized by PrOF NMR to better understand their activity in solution. Notably, due to the importance of the disordered protein α-synuclein in neurodegenerative diseases, Pielak and colleagues have characterized its conformational changes and binding interactions also using PrOF NMR.[90-92]

Kinetic information regarding protein folding, denaturation, and refolding is also acquirable through the use of PrOF NMR which not only demonstrates its versatility but also its speed.[93] Frieden and colleagues reported the use of stopped-flow ^{19}F NMR and

two different proteins (*E. coli* Dihydrofolate Reductase and PapD) labeled with 6-fluoro-tryptophan (6FW)[93-96] or 4-fluoro-phenylalanine (4FF)[97] to monitor protein folding in solution under various conditions. By observing the relationship between peak area and mixing time, they could better understand the dynamics of proteins in solution as they fold in their native state, denature under specific conditions, and refold when denaturing conditions have been removed.[95] Some of their spectra were acquired as fast as 1.5[93] and 2.74 seconds[97] demonstrating the speed of PrOF NMR.

PrOF NMR has been used to characterize ligand binding and to screen libraries of compounds in multiple cases in recent years. Gee et al. screened 508 fragments against the transcriptional coactivator KIX and were able to identify several new small molecule ligands for KIX. These results and their follow-up will be further discussed in chapters 2 and 3. PrOF NMR has also been used to characterize ligand binding to a number of bromodomains,[98] was used in a dual screening format to identify a selective bromodomain ligand,[99] and been compared to ^{1}H-CPMG NMR, a standard industry NMR screening technique.[75]

1.2.4 Fluorine Side-Chain Resonance Assignment

Once a protein has been labeled with fluorinated amino acids, it is important to determine which resonances correspond to which amino acids. Several methods to accomplish this goal are used, including site directed mutagenesis,[41] site-specific labeling,[100] ligand binding experiments, and solvent exposure experiments with broadening agents[41] as well as multidimensional NMR experiments.[101-102] Further descriptions, procedures, and anticipated results for site directed mutagenesis will be discussed in Chapter 4.

The use of several sequential multidimensional NMR techniques for assigning the spectrum of ^{19}F-labeled side-chain resonances in proteins was pioneered by the lab of Scott Prosser.[101] These techniques seek to avoid the challenges inherent in mutagenesis approaches such as difficulty of expression and spectral overlap. Prosser's method utilizes several through-bond correlations to link the fluorine resonances to their corresponding amide nitrogen and consequently their position within the primary sequence.

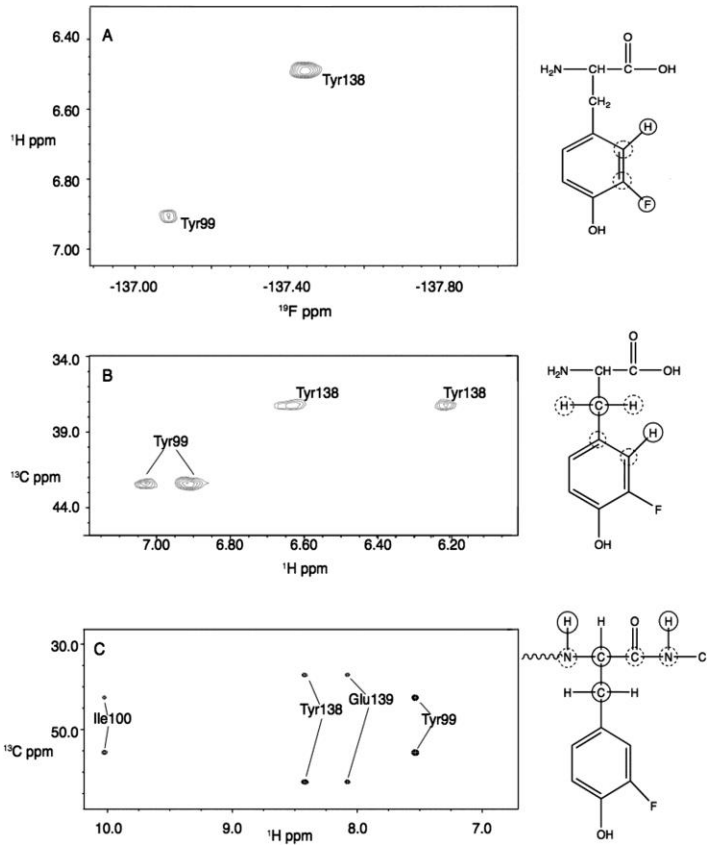


Figure 1.10 Example spectra demonstrating the use of multidimensional NMR experiments to assign fluorine resonances.
A) F-H COSY correlating the fluorine and its neighboring proton. B) Correlation of aromatic (δ) and β protons. C) HNCACB to correlating the β proton to the amide backbone. Reprinted with permission from Kitevski-LeBlanc et al.[101] Copyright 2009 American Chemical Society.

Using 3-fluorotyrosine-labeled calmodulin (CaM) as the example, the first experiment uses a CT-HCCF-COSY pulse sequence which is a slightly modified COSY experiment to correlate the fluorine to its neighboring proton on the aromatic ring.[101] The second experiment correlates that proton to the ones on the β carbon using an (Hβ)Cβ(CγCδ)Hδ pulse sequence. Lastly, the β protons are correlated to the amide backbone using a standard HNCACB experiment. Through this method, each fluorine resonance can be assigned to its corresponding amino acid position based on the correlation data obtained through multidimensional NMR.[101] A similar method has also been used for the assignment of 3-fluorophenylalanine-labeled CaM.[102]

While this method does remove the challenges of the mutagenesis approaches, this multidimensional technique requires the protein to be labeled with ^{15}N and ^{13}C in addition to the fluorinated aromatic amino acids. Furthermore, this method is only effective if the ^{1}H-^{15}N HSQC assignments have already been obtained. It does, however, provide an alternative when mutagenesis approaches are unsuccessful.

As an additional alternative to site-directed mutagenesis, Isley et al. demonstrated the possibility of computationally predicting the chemical shifts of fluorine resonances in a given protein based on its environment, orientation, neighboring water molecules, and a range of other factors.[103] While this method is not yet accurate enough to perfectly predict chemical shifts of each resonance, it is a proof-of-concept that chemical shifts of fluorinated amino acids in proteins may be able to be predicted in the future.

1.2.5 Paramagnetic Metals and Relaxation Times

The transition from 2D or 3D NMR experiments to a 1D ^{19}F experiment significantly reduces the total experimental time, but there are still two limitations: relaxation time and chemical shift dispersion. Even short 1D experiments are limited by the delay time in the pulse sequence which is typically 3-5 times the longitudinal (spin-lattice) relaxation time (T_1).[53] In addition, NMR data interpretation can be hindered if there are overlapping or poorly resolved peaks. Both of these challenges can be addressed through the addition of paramagnetic metals, most commonly the paramagnetic lanthanides.[54]

Paramagnetic compounds such as O_2 and certain metals, such as Ni^{2+}, Dy^{3+}, Ho^{3+}, Cu^{2+} and Tm^{3+}, have a profound effect on the relaxation of nuclei in magnetic resonance experiments.[70, 104-106] Therefore, paramagnetic agents can be used to intentionally modulate relaxation properties as described below and in Chapter 4.[107] Hawk et al. demonstrated that the addition of 20 mM Ni-DTPA could shorten T_1 relaxation times by approximately 75%, thereby shortening the overall experiment time by approximately 60%. The inverse principle is also useful i.e., monitoring changes in relaxation to determine the amount of paramagnetic agent present, as we reported in ACS Nano and as described in Chapter 6.[108] Historically lanthanides have been used as Magnetic Resonance Imaging (MRI) contrast agents. By shortening relaxation times, the contrast in

the MRI readout improves.[109] It is also well known that lanthanides are capable of changing the chemical shifts of different peaks in an NMR spectrum such that formerly overlapping peaks can be resolved.[110] The degree to which each metal alters the chemical shift and the relaxation times differs and is portrayed in Figure 1.11.[104]

The use of paramagnetic metals in shortening relaxation times applies not only to small molecules but also has literature precedence to aid in protein NMR. Ramamoorthy and colleagues have frequently utilized the paramagnetic properties of the Cu^{2+} to shorten relaxation times in their studies of protein structure and function.[111-114] They have observed a 3-10-fold decrease in T_1 relaxation with no significant change in T_2 relaxation at concentrations as low as 3 mM Cu^{2+}.[111, 113] Similarly Ishii et al. reported a 4 fold decrease in T_1 relaxation with the addition of Cu^{2+}.[115]

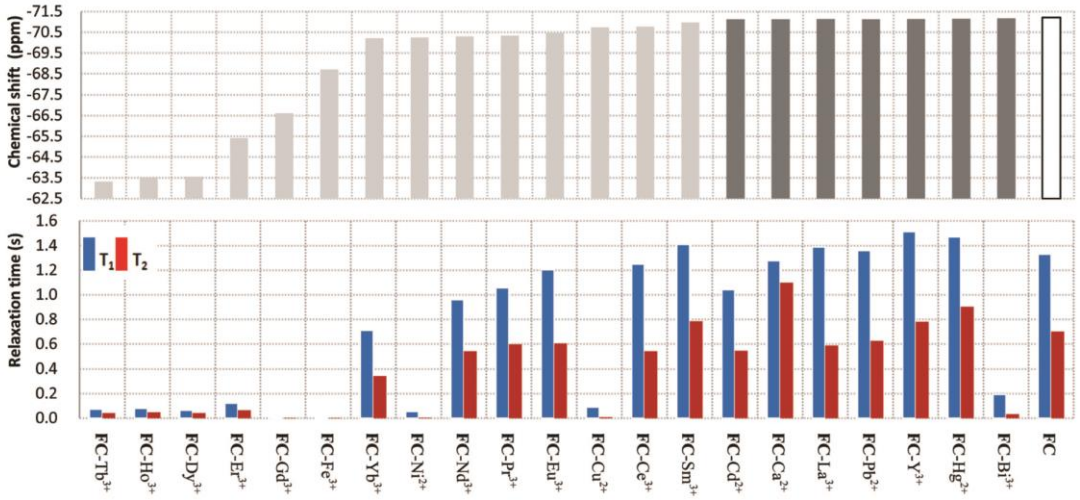


Figure 1.11 Effect of various metal ions on ^{19}F NMR signals.
(A) Chemical shift in the presence of metal ions (B) Impact on T_1 and T_2 relaxation times. Adapted from Jiang et al.[104] with permission of the Royal Society of Chemistry.

Both the T_1 and the T_2 relaxation times are important to understand and play a significant role in data acquisition. The T_1 relaxation time, also called the spin lattice or the longitudinal relaxation time, is qualitatively described as the amount of time required for a signal to precess back to its original position (M_0) in the Z axis after being radiated with the radio frequency (RF) pulse. The faster the nuclei return to their original

magnetization, the sooner they can be pulsed again, meaning the experiment can be completed in a shorter time.[53]

The T_2 relaxation time, also known as the spin-spin or transverse relaxation time, describes the amount of time required for the nuclei to lose phase coherence.[53] T_2 can also be thought of as the time required for the net magnetization vector in the XY plane to return to zero.[116] As the nuclei precess back to their original position in the Z-axis, they do so in a circular manner around the Z axis. As they precess around the Z axis, the XY magnetization slowly starts to lose phase coherence which results in a broader signal.[53] Ideally when altering the relaxation properties of a given compound, the change would balance the shortest T_1 possible with the longest T_2 possible.

The relaxation process is heavily dependent on its surrounding environment and is carried out through a variety of mechanisms including dipole-dipole, chemical shift anisotropy (CSA), spin-rotation, and quadrupolar relaxation. Generally the most significant of these four is the dipole-dipole interaction in which the dipole from a nearby nuclei interacts with the nuclei of interest.[53] According to the Solomon-Bloembergen equations, the magnitude of the effect of the dipole-dipole interaction is directly proportional to the magnetic moments created by both spinning nuclei and inversely proportional to the distance between the two nuclei in an r^{-6} manner.[117] Unpaired electrons, such as those in paramagnetic metals, can generate a magnetic moment that is over 600 times the magnetic moment of a neighboring proton which results in an even more significant reduction in relaxation time. This paramagnetic relaxation effect can occur at distances up to 35 Å away in some cases.[53, 110, 118] Because the T_1 and T_2 relaxation times are described by different equations (Eq 3 and 4) as noted by Solomon[117] and Bloembergen,[119] a reduction in one will not result in an identical reduction in the other.

$$T_1 = \frac{2}{5}\left(\frac{\mu_0}{4\pi}\right)^2 {\gamma_1}^2 g^2 {\mu_B}^2 S(S+1) J_{SB}(\omega_1) \quad (3)$$

$$T_2 = \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 {\gamma_1}^2 g^2 {\mu_B}^2 S(S+1)[4J_{SB}(0) + 3J_{SB}(\omega_1)] \quad (4)$$

where μ_0 is the permeability of free space, γ_1 is the gyromagnetic ratio, g is the electron g-factor, μ_B is the magnetic moment of the free electron, $\omega_1/2\pi$ is the Larmor frequency and J_{SB} is described as follows:

$$J_{SB} = r^{-6} \frac{\tau_c}{1+(\omega\tau_c)^2} \quad (5)$$

where τ_c is the correlation time.[118]

To measure the T_1 relaxation time for a given compound, the inversion recovery (IR) experiment is used. In this NMR experiment (Figure 1.12), the time delay, denoted by τ, is arrayed from small to large values. At a shorter time, τ, the magnetization of the nuclei has not adequately relaxed back to its equilibrium position (M_0) resulting in a reduced signal following the 90° pulse. As τ increases, the signal intensities of the peaks in the NMR spectrum also increase. The integrals of the peaks from the array are then fit to an exponential equation to determine the T_1 (Eq. 6).[53, 71, 116]

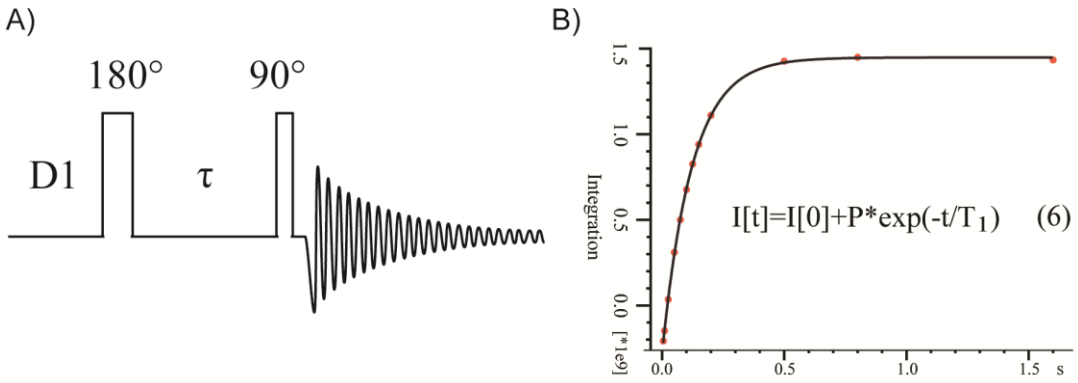


Figure 1.12 T1 relaxation measurements.
A) NMR pulse sequence with a 180° pulse followed by a 90° pulse with a variable delay time (vd). B) Sample T_1 data and corresponding equation to fit the data.

The T_2 relaxation time can be evaluated by two different methods. An approximation utilizes the relationship between line broadening and T_2 as described in the following equation:

$$T_2 = \frac{1}{\pi * LW} \quad (7)$$

where LW is equal to the linewidth at half height. This approximation assumes perfect homogeneity within the NMR sample which is often times not the case.[53, 71, 116] Because that method is only an approximation, other techniques were developed to more

accurately measure the T_2. The first pulse sequence for T_2 determination was developed by Carr and Purcell[120] but was later modified by Meiboom and Gill, yielding the now standard Carr-Purcell-Meiboom-Gill (CPMG) experiment.[121] This method provides a more accurate measurement of the T_2 relaxation time by examining a data array in a similar manner to the T_1 IR experiment. Delay times, or number of repetitions of the cycle, are arrayed such that at a shorter delay there is a larger signal and at the longer delay the signal is significantly reduced or nonexistent. The data can be fit using a first order exponential decay equation. (Figure 1.13)

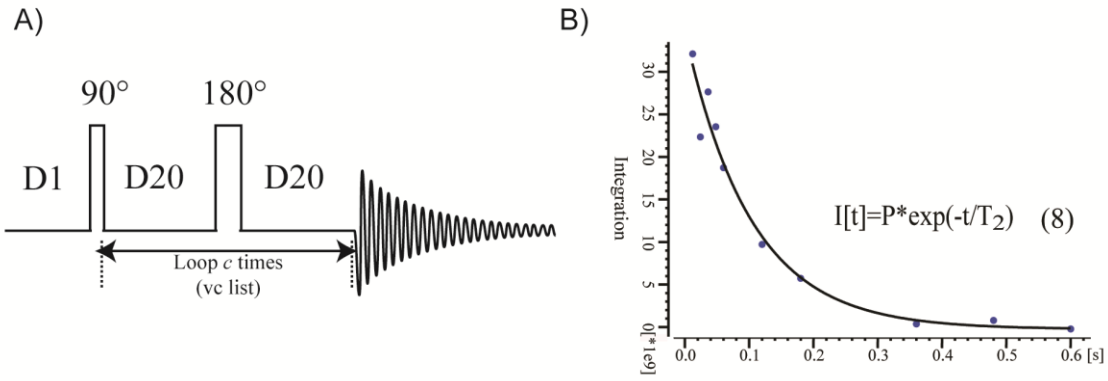


Figure 1.13 T2 relaxation measurements.
A) NMR pulse sequence with a 90° pulse followed by a 180° pulse with a set echo time (D20). The D20-180° pulse-D20 sequence will be repeated *c* times. Total experiment time will be determined by the vc list. B) Sample T_2 data and corresponding equation to fit the data.

The lanthanide induced shift (LIS) effect arises from the paramagnetic property known as pseudocontact shift (PCS) in which the magnetic field generated by unpaired electrons affects the nuclei of interest in a r^{-3} distance dependent manner.[118, 122-123] This effect is only possible in an anisotropic environment, meaning there is an uneven distribution of electrons in the magnetic field.[124] The necessity of anisotropy for LIS stands in contrast with the paramagnetic relaxation effect which can be induced with either isotropic or anisotropic magnetic fields. For this reason not all paramagnetic metals can be used as shift reagents.

1.2.6 NMR physics of spin relaxation and molecular motion

Nuclear magnetic resonance (NMR) spectroscopy experiments are dependent on the relaxation properties of the nuclei being investigated. Nuclear spin relaxation is a process by which a nuclear spin interacts with any external magnetic field, whether that is the magnetic field from the spectrometer, neighboring nuclei, or free unpaired electrons. When the nuclear spin of interest is perturbed, it relaxes back to its equilibrium spin state in accordance with its own magnetic properties and the properties imparted to it. Some examples of these interactions include dipolar coupling, quadrupolar interactions, spin-rotation, and chemical shift anisotropy (CSA).[53] One commonality among all of these relaxation mechanisms is their dependence on molecular motion which is described by various spectral density functions. CSA relaxation can have a dominating effect on ^{19}F NMR spectra, so it is important to understand its origins and implications.[125]

Chemical shift or chemical shielding anisotropy (CSA) in its simplest conceptual description arises from asymmetric or anisotropic dispersion of electrons in a given system. Nuclei with more electrons, like ^{19}F, are anisotropic and more susceptible to effects from CSA than nuclei that have fewer electrons, like ^{1}H which are isotropic.[53] Molecular tumbling in solution allows local magnetic fields to be generated which aid in spin relaxation. The magnitude of the effect of these generated fields will depend on the degree of anisotropy present in the nucleus of interest.

The most noticeable effect on an NMR spectrum from CSA is line broadening due to the dependence of T_2 relaxation on the square of both the shielding anisotropy value

$$\frac{1}{T_2} = \frac{2}{15}(\sigma_{\parallel} - \sigma_{\perp})^2 {\omega_F}^2 \tau_c \left[\frac{2}{3} + \frac{1}{2}\frac{1}{(1+\omega_F{}^2\tau_c{}^2)}\right] \quad (9)$$

where T_2 is the spin-spin relaxation time, $(\sigma_{\parallel} - \sigma_{\perp})$ is the CSA, ω_F is the strength of the magnetic field, and τ_c is the rotational correlation time.[125] While CSA for any given nucleus will vary depending on its chemical environment such as nearby functional groups or position within a protein, the values for each nucleus tend to fall within predictable ranges. Since ^{1}H is mostly isotropic, its CSA is typically between 0 and 20 ppm. The CSA for ^{13}C tends to fall between 80 and 120 ppm but can be as low as 15 ppm.[126] The CSA for ^{15}N varies significantly depending on its environment, but in the

peptide backbone of a protein, it tends be between 160 and 180 ppm.[127] ^{19}F also has a large CSA (50-100 ppm).[128-130] In a direct detection experiment of these nuclei, ^{15}N resonances would be the broadest on the basis of CSA. This effect is partially mitigated and some signal sensitivity is recovered, however, by the fact that for an HSQC, ^{15}N is indirectly detected through the ^{1}H nucleus. Consequently, the ^{15}N resonances retain the signal sensitivity and some sharpness from ^{1}H due to dipolar relaxation and low CSA.[53] With the advent of transverse relaxation-optimized spectroscopy (TROSY) NMR methods, the CSA challenge for ^{15}N labeled proteins is surmountable.[58] In contrast, there are currently no TROSY methods available for direct-detected ^{19}F experiments, and therefore nuclei experience the full effects of CSA.

The T_2 relaxation is also dependent on the square of the magnetic field strength. (Equation 7) Therefore, nuclei whose relaxation mechanisms are dominated by CSA experience an even greater broadening effect at higher magnetic fields (Figure 1.14).[131] Consequently, these nuclei can be more efficiently analyzed at lower magnetic fields (e.g. 500 MHz instead of 900 MHz). The rotational correlation time (τ_c) can further compound the broadening of linewidths as an increased τ_c also results in increased linewidths, as is the case in a protein backbone. However, internal rotational motion within a protein side chain can lead to line narrowing, allowing for the possibility of analyzing large proteins by NMR.[63]

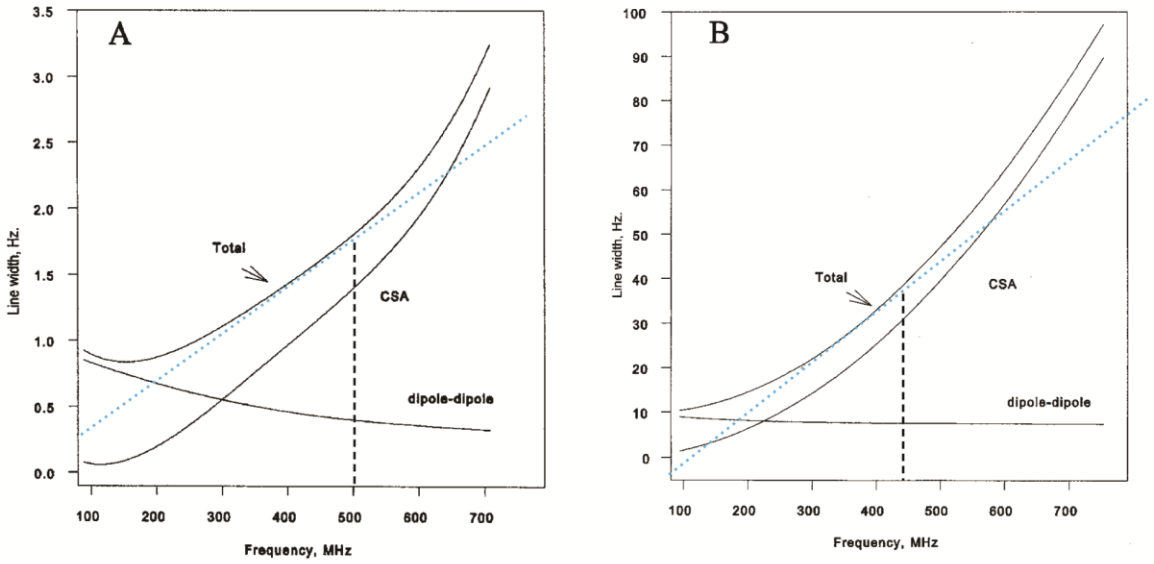


Figure 1.14 Predicted effect of CSA and dipole-dipole relaxation on ^{19}F linewidths at varying magnetic field strengths.
A) Rotational correlation was 0.4 ns B) Rotational correlation time was 16 ns. ····· Extrapolation of the linear region of CSA. - - - - - Optimal magnetic field strength before CSA effects take over. Adapted from Gerig et al.[131]

In its basic mathematical form, chemical shift anisotropy, is a tensor, which means it is calculated relative to three components in a coordinate axis system. The chemical shift tensor is denoted as σ while its component parts are designated σ_{11}, σ_{22}, and σ_{33}.[125] For an isotropic system like ^{1}H, these values are equivalent (Equation 10).

$$\sigma_{iso} = \frac{1}{3}[\sigma_{11} + \sigma_{22} + \sigma_{33}] \quad (10)$$

For systems that deviate from the ideal isotropic sphere, the anisotropy can be calculated using equation 11 and abbreviated as equation 12.

$$\Delta\sigma = \sigma_{11} - \frac{1}{2}(\sigma_{22} + \sigma_{33}) \quad (11)$$

$$\Delta\sigma = \sigma_{\parallel} - \sigma_{\perp} \quad (12)$$

using the standard convention that $\sigma_{11} \geq \sigma_{22} \geq \sigma_{33}$.[129]

These tensor components can be predicted through computational studies using quantum mechanics, but they also can be measured or calculated from experimental data.[126, 128] Using solid state NMR, measurements can be taken on single crystal samples, static powder pattern spectra from polycrystalline samples, and spinning side bands undergoing magic angle spinning (MAS), in addition to CSA recovery experiments that

measure CSA when it is normally cancelled out.[132] In solution NMR, the CSA tensors can be calculated based on examining T_1 and T_2 relaxation data provided that CSA is the predominant relaxation mechanism relative to dipolar interactions. Cross correlated spin relaxation between CSA and dipole-dipole (DD) relaxation can also provide information regarding the tensor values.[133-134]

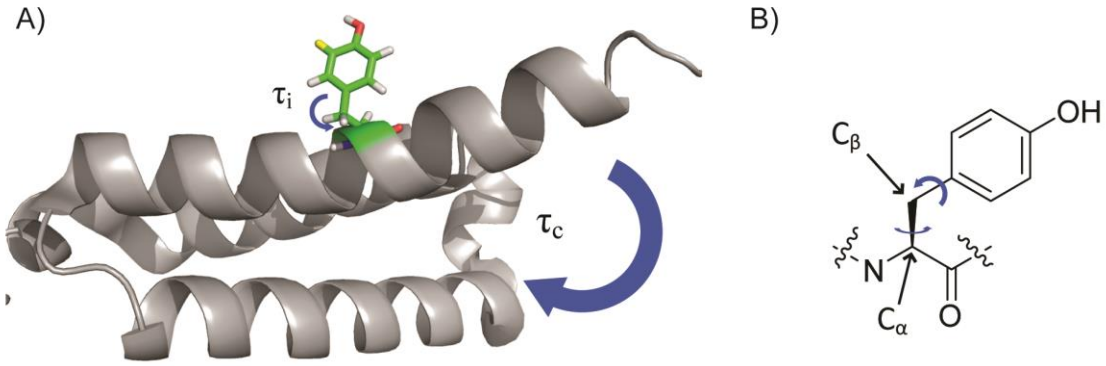


Figure 1.15 Molecular motion of a protein.
A) The overall tumbling of a protein in solution is represented by τ_c while the internal rotation of side chains is represented by τ_i. B) Internal side-chain rotation occurs around the C_α-C_β bond and the C_β-aromatic bond.

In addition to CSA, another key contributor to resonance linewidths is molecular motion in solution.[131] In the case of proteins both the overall tumbling in solution and the internal side-chain rotation must be taken into consideration (Figure 1.15 A). The rotational correlation time, τ_c, is a function of molecular weight such that larger molecules tumble slower in solution, leading to line broadening. Internal correlation time τ_i corresponds to internal side chain rotational motion which most often leads to line narrowing, as calculated using spectral density functions.[125]

For aromatic amino acids, the dominant source of internal rotation is around the C_β-aromatic bond (Figure 1.15 B). Some internal rotation also occurs around the C_α-C_β bond. In the case of phenylalanine, this effect can result in narrower linewidths due to τ_i being faster than τ_c. Tryptophan, due to its large size, and tyrosine to an extent, have very limited rotation around this bond when they are buried within a protein. In these cases τ_i can be approximated to be equal to the τ_c of the protein as a whole.[135-137] This

phenomenon was estimated by Hull and Sykes[125] (Table 1.2) and later calculated based on experimental data by Dürr et al.[129] (

Table **1.3**). This same approximation cannot be made, however, when the residues are more surface or solvent exposed.

Table 1.2 Predicted ^{19}F linewidths of 3FY and 4FF at varying field strengths, molecular weights, and internal side chain rotational correlation times
Linewidths are reported in Hz, and 7 x 10^{-8} s was used for the overall rotational correlation time, τc. "Isotropic" is used to denote no internal motion. Adapted from Hull and Sykes[125] with permission from Elsevier.

NMR Frequency	log τ_i	3-fluoro-tyrosine		4-fluoro-phenylalanine	
		MW=86,000	MW=20,000	MW=86,000	MW=20,000
94	-11	3.2	0.46	0.07	0.01
	-10	3.2	0.46	0.08	0.01
	-9	3.2	0.50	0.11	0.04
	-8	3.4	0.69	0.34	0.24
	-7	4.8	1.1	1.7	0.67
	-6	7.7	1.3	4.9	0.88
	Isotropic	8.8	1.3	6.3	0.92
254	-11	23.0	3.3	0.5	0.08
	-10	23.0	3.3	0.6	0.11
	-9	23.3	3.6	0.8	0.34
	-8	24.8	4.7	2.2	1.5
	-7	35.0	7.6	12.5	4.7
	-6	55.8	9.0	35.9	6.3
	Isotropic	64.3	9.2	45.8	6.6
339	-11	40.9	5.9	1.0	0.14
	-10	41.0	5.9	1.0	0.19
	-9	41.4	6.3	1.4	0.58
	-8	44.1	8.3	3.8	2.5
	-7	62.2	13.5	22.2	8.4
	-6	99.2	16.0	63.9	11.2
	Isotropic	114.2	16.3	81.5	11.7

The linewidths in Table 1.2 were calculated using reported tensor values for fluorobenzene compounds, with monofluorobenzene and 1,2-difluorobenzene serving as approximations for 4-fluorophenylalanine (4FF) and 3-fluorotyrosine (3FY) respectively.[125] As expected, both amino acids show an increase in linewidth as a function of both increasing field strength and increasing molecular weight, which, as mentioned above, relates to the rotational correlation time of the protein as a whole. For

both amino acids, there is also significant line broadening as the amount of internal side-chain motion decreases. For 3FY the linewidth experiences a 3-4 fold increase from comparing fast internal rotation to no internal rotation. For 4FF, this effect was even more pronounced where the linewidths increased by almost two orders of magnitude when the residue was calculated to have no internal motion. In comparing the two amino acids, 4FF has significantly narrower linewidths than 3FY, which serves as evidence that 4FF labeling of a protein may provide the narrowest resonances.

Table 1.3 CSA and relaxation data obtained by solid state ^{19}F NMR, Adapted from Dürr et al.[6]

Amino Acid	CSA (ppm)	T_2 (μs)
4FF	60.5	40.6 ± 7.2
3FY	74.6	25.4 ± 5.4
5FW	47.1	42.0 ± 1.1

Using solid-state NMR data of polycrystalline amino acids, Dürr et al compared the CSA and relaxation times of various fluorinated amino acids that have been incorporated into poly peptides and proteins.[129] Key results have been summarized in

Table **1.3**. This data corroborates the predictions made by Hull and Sykes with 3FY being more affected by CSA and consequently having as shorter T_2 relaxation time than 4FF. Given the relationship between T_2 and linewidth, it is also reasonable to postulate that the 3FY resonances will be broader than those of 4FF. 5FW has the lowest CSA and the longest T_2 of the three amino acids being compared which would make it an ideal candidate for studying a protein. However, as established above, the internal side-chain rotational motion must also be taken into consideration. Both 4FF and 5FW are comparable in regards to their T_2 relaxation, but their internal motion will be heavily dependent on their location within a protein in regards to whether it is more surface exposed or buried in the interior of the protein.

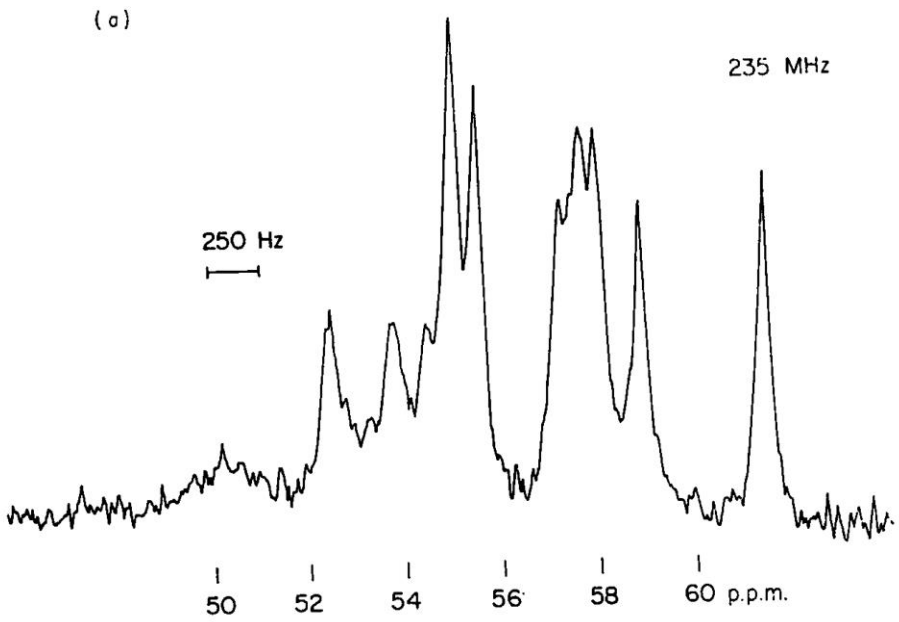


Figure 1.16 3FY labeled alkaline phosphatase obtained on a 250 mHz NMR. Fluorine incorporation was approximately 70%. Adapted from Hull and Sykes[125] with permission from Elsevier.

When considering linewidths in NMR, it is important to understand the many components that contribute, including the chemical shift anisotropy of the nucleus being, the rotational correlation time of the protein as a whole, and the internal side-chain rotation. While ^{19}F may have more CSA than ^{15}N, by labeling side chains rather than the backbone it is still possible to obtain distinct resonances even at large molecular weights. Even though 3FY is expected to be the least favorable aromatic amino acid to be incorporated into a protein due to broadening and poor mobility, it is still feasible to discern the individual residues on alkaline phosphatase (MW 86 kDa) (Figure 1.16). While some resonances overlap, most are distinguishable. To summarize, even though CSA has significant effect on line broadening, PrOF NMR is still well suited to study proteins as demonstrated by its use for various proteins encompassing a wide range of molecular weights, its sensitivity to its chemical environment, and its ease of use and interpretation.[41, 79, 90, 138-140]

1.3 **Small Molecule Screening**

The identification of viable compounds for drug use is a difficult process that has been addressed by screening large libraries of compounds against a target of interest. To efficiently select potential drug compounds, Lipinski's "Rule of 5" (RO5) is often used as a guideline.[141] The RO5, as well as other iterations like the Rule of 3 (RO3)[142] provide guidelines regarding size and chemical properties for selecting compounds that will have a greater likelihood of success as drug compounds. High throughput screening (HTS) has been the dominant method for drug discovery, but recent years have also seen the rise of fragment based screening (FBS) as an alternative and potentially more efficient method.[143-144]

1.3.1 High Throughput Screening

HTS is characterized by the ability to screen more than 10,000 compounds within one week, although with advances in technology it is not unreasonable to now expect 10,000 compounds per day. By nature of the screening method, any analysis technique for HTS must be extremely fast. This lends itself best to optical methods which can be examined quickly with a camera or other optical detector. These techniques include biochemical assays such as fluorescence polarization and cell based assays such as reporter gene assays. The application determines both the types of compounds that are screened as well as the technique used for screening.[145-146]

While HTS has been well optimized, there are intrinsic shortcomings that it has yet to address. One major challenge is the size of chemical space, which is a description of the number of molecules possible given certain conditions. In HTS, chemical space is estimated to be composed of 10^{60} unique molecules.[147] Even screening at a rate of 200,000 compounds per day for a year does not come close to covering all of chemical space. While HTS has had many success stories for a variety of protein classes, most notably enzymes, the challenge of low hit rates for PPIs is sufficient cause for seeking a more efficient method for small molecule discovery.[143] HTS faces a qualitative challenge as well in that it seeks to identify high affinity compounds ($K_d \leq 1\mu M$).[147] The number of lead compounds with this level of affinity is extremely low, which adds to the problem of already not covering very much of the available chemical space. These challenges may

indicate that the types of compounds in HTS libraries are not fit for targeting PPIs. If there was a way to simultaneously reduce the size of chemical space and increase the hit rate, the efficiency of the entire process would be greatly improved.

1.3.2 Fragment Based Screening

Fragment based screening (FBS), developed as an alternative to HTS, differs in the types of molecules that are screened. FBS examines low molecular-weight, low affinity fragments of drug-like molecules rather than full size drug molecules. A hit in FBS typically has a K_d in the mid μM to low mM range.[147] These fragments that bind in nearby locations in the protein can then be pieced together to form active drug compounds possessing higher affinities than any of the component fragments.[148] In addition to a higher hit rate, chemical space in FBS is significantly smaller than it is in HTS. The number of potential fragment molecules in chemical space for molecules less than 160 Da is estimated to be on the order of 10^7 in contrast with 10^{60} as is estimated for complete chemical space in HTS.[147] This drastic reduction in the size of chemical space makes exploring the entirety of chemical space far more feasible.

NMR is an ideal FBS technique and holds numerous advantages over other methods. It requires no immobilization or bulky fluorophores. It is also a non-destructive technique, allowing the protein to be reused. Additionally, NMR is performed in a buffered solution at a physiological pH which allows the proteins to orient and fold naturally as it would in a biological system. It also provides structural information regarding the protein of interest. Sample preparation is also significantly easier in NMR than it is for most other screening techniques. Using SAR by NMR, as discussed earlier, Fesik and colleagues have been able to discover small molecule fragments that bind to proteins. As a proof-of-concept, Pomerantz et al. used PrOF NMR for a preliminary fragment screen of 50 small-molecule fragments which recognized one hit as determined by a $\Delta\delta$ of 0.100 ppm in the ^{19}F NMR spectrum. The conclusion was that not only can PrOF NMR be used for FBS but that it could be done within a reasonable amount of time (5-10 min/fragment mixture).[41]

1.3.3 Screening Metrics for Ligand Development

In a screening campaign, there are a range of metrics that can be used to prioritize which compounds should be carried on to development and which should be triaged. Affinity is just one of many metrics that is utilized to evaluate hit compounds from a screen. Beyond principles avoiding pan-assay interfering compounds (PAINS)[149] or identifying compounds that will be easier to synthetically derivatize, there are several versions of what has become commonly known as ligand efficiency. Here we will describe just two of them.

A compound's ligand efficiency (LE) is defined as the its change in free energy per non-hydrogen atom. Essentially it is a measure of how much each atom in the molecule contributes to the overall binding energy and is calculated using Equation 13 where HA denotes the number of heavy or non-hydrogen atoms. This equation can be modified to be calculated from a K_d, as shown in Equation 14 where R is the ideal gas constant, T is the temperature in Kelvin, and K_d is the dissociation constant.

$$LE = \Delta G/HA \tag{13}$$

$$LE = -RT \ln K_d/HA \tag{14}$$

Each atom in a compound will either contribute to or take away from the binding energy of the compound to the protein target. Therefore it is advantageous to start with a compound that has a high LE as opposed to a low LE because a low LE makes the development process more difficult. In a case study of 480 target-assay pairs, the median LE was found to be 0.32 kcal mol^{-1} HA^{-1} with a range of 0.2 to 0.6 kcal mol^{-1} HA^{-1}.[150]

A compound's lipophilic ligand efficiency (LLE) is calculated using equation 15 and describes the specificity of a compound to bind to the target relative to its affinity to partition into 1-octanol. Note that in this case the LLE can be calculated using any variation of an affinity measurement (e.g., K_i, IC_{50}, EC_{50}).

$$LLE = pK_d - clogP \tag{15}$$

The LLE accounts for not only affinity but also for hydrophobicity. It is not uncommon for developed compounds to be triaged due to unfavorable adsorption, distribution, metabolism, excretion, and toxicity (ADMET) properties.[151] Analyzing a hit's LLE is a method to preemptively avoid these types of issues later on since lipophilicity is often

closely related to other ADMET properties. In the same study cited above, the median LLE for the 480 target-assay pairs was 2.83 with a range from -3 to 9.[150] An ideal drug candidate is estimated to have an LLE greater than 5, but it is common for HTS hits to have an LLE $\leq$ 2. Further discussions of both LE and LLE will be presented in chapters 2 and 5.

1.4 **Additional Techniques for Ligand Screening**

1.4.1 NMR spectroscopy

While protein-observed NMR techniques have been discussed at length above, it is worth noting that ligand-observed techniques such as saturation transfer difference (STD), Water-Ligand Observed via Gradient Spectroscopy (WaterLOGSY), and Carr-Purcell Meiboom-Gill (CPMG) are standard screening tools utilized by industry and pharmaceutical companies.[144, 152-154] These ligand-based techniques have the distinct advantage of removing the need for a deconvolution step when screening mixtures of compounds. Because resonances corresponding to specific ligands are monitored, it is clear which ligands are undergoing binding events and which are unaffected.

Urick et al. performed a comparative screen using PrOF NMR and ^{1}H CPMG NMR against the protein Brd4.[75] They compared the hit rates, speed, material consumption, propensity for false results (positive or negative), the ability of each method to triage compounds, and the responsiveness of the each method. Results were confirmed by thermal shift assays (Section 1.4.2) and ITC. Urick et. al. reported an 80% overlap between the two assays, demonstrating the complementary nature of the two techniques.

It has also been common to use ^{19}F NMR to perform ligand observed experiments in the cases where libraries of fluorinated compounds are available.[155-156] Much of this work has been pioneered by Dalvit and coworkers over the past 15 years in performing both 1D and 2D ^{19}F NMR experiments for small molecule discovery. Of relevance is the fact that almost 30% of all FDA approved drugs contain at least one fluorine, making the screening of fluorine-containing small-molecule libraries even more attractive.[157]

1.4.2 Thermal Shift Assays

Thermal shift assays, such as differential scanning fluorimetry (DSF), are another class of biophysical assays commonly used as screening platforms.[158-160] Their rapid throughput and low material requirements make them an attractive starting point for a screening campaign. The premise of a thermal shift assay is that as a ligand associates with a protein of interest, it will stabilize or destabilize the structure of the protein, thereby changing or shifting the melting point (T_m) of the protein. Conversely compounds that do not bind to the protein will have no effect on the melting point of the protein. Typically a fluorescence of a dye is monitored as a function of temperature. As the temperature increases and the protein begins to unfold, hydrophobic regions of the protein are then exposed, to which the fluorescent dye binds. This association increases the fluorescence signal that is measured. The more stabilized a protein is by a ligand, the higher the temperature will need to be to achieve the same fluorescence signal.

Molina and coworkers. have reported on methods for a cellular thermal shift assay (CETSA) in which the same principles discussed above are performed but in a cellular context.[161-162] The consideration of the cellular environment is important because a drug compound may have a different on- or off-target effect in cells than it does in an in vitro system. This method has significant potential to help triage compounds in the pre-clinical drug development stage and further aid in the process of studying drug efficacy at biologically relevant levels.

Thermal shift assays provide incredibly rapid screening results, but its main limitation is in its inability to provide any quantitative or structural information about the newly discovered ligand. Loose correlations can be drawn between the magnitude of change in a protein's melting temperature and the affinity of a given compound. There is also a significant amount of debate regarding the relevance of destabilizing compounds i.e., compounds that decrease rather than increase the T_m. While most people tend to favor compounds that stabilize a protein target, Gestwicki and coworkers have demonstrated uses for destabilizing compounds as well.[163]

1.4.3 Surface Plasmon Resonance

SPR is one of the more common screening techniques available due to its rapid throughput and low material consumption. [164] In SPR, a protein of interest is immobilized on the surface of a chip, and ligands are flowed over the protein, often using a microfluidics platform. On the other side of the surface to which the protein is attached is a layer of gold. This layer is excited by plane polarized light, which releases a plasmon wave. The frequency of this wave will change as a function of changes on the surface, such as ligand binding to a protein. The altered plasmon wave also changes the light that is reflected from the surface. Consequently, binding events can be monitored in real time, allowing the measurements of on-rates (K_{on}) and off-rates (K_{off}).

One of the key advantages of SPR is that it is a label-free technique. However, the protein does have to be immobilized on a surface which can still potentially alter ligand binding, either positively or negatively. This consideration further emphasizes the need for orthogonal methods of evaluation of hit compounds.

1.4.4 Virtual Screening

A relatively recent tool that has been gaining popularity and utility is virtual screening, which computationally predicts ligand binding.[165] A range of softwares and methods are available, but they all operate under similar principles. At a basic level, these "docking" studies will take a molecular mechanics approach to analyzing whether or not a small molecule will fit into a binding pocket. Its fit will be evaluated on criteria such as steric hindrance and charge interactions. More advanced methods utilize quantum mechanics for more precise calculations. Because virtual screening is a computational method, the throughput and capacity of compounds that can be screened is unparalleled in biophysical assays. We reported on SAR development that came from a six million compound virtual screen against Brd4.[166] This work is also highlighted in Appendix 4.

The main drawback to virtual screening is the fact that it is virtual. As good as computational predictions are, they do not perfectly reflect what happens in a physical experiment. As such, it is a very useful tool for initial screening and for brainstorming new compounds and analogs, but it must be followed up with biophysical assays.

1.5 **Brief descriptions of chapters**

The following chapters describe my role in various projects in the development and application of fluorine-based magnetic resonance experiments. Chapters 2 and 3 focus on the application of PrOF NMR to the protein KIX in a full fragment screen (Ch. 2) and in the discovery of a new ligand binding site (Ch. 3). Chapter 4 is a description of our PrOF NMR protocol, covering the entire process from bacterial transformation to fragment screening. Chapter 5 shifts in focus to bromodomains, a class of proteins that are biologically of interest as epigenetic "reader" proteins. This chapter encompasses three separate projects for targeting the bromodomain BrdT from a ligand deconstruction approach and from a fragment screening approach. Chapter 6 describes the application of ^{19}F NMR and MRI as an oximetry tool using ultraporous mesostructured silica nanoparticles loaded with perfluorocarbons. Finally, Chapter 7 features an educational report on protein quantification using a technique known as smartphone colorimetry, making experiments like the Bradford assay accessible even in situations when a UV-Vis spectrophotometer is not available. Appendix 1 is an extension of Chapter 6, describing more of the details in the synthesis and characterization of the nanoparticles. Appendix 2 features both a student worksheet and instructor guide for the experiments outlined in Chapter 7.

Chapter 2. Fragment Screening and Druggability Assessment for the CBP/p300 KIX Domain Via Protein Observed ^{19}F NMR

Motivation: The motivation for this manuscript was to address the question of whether or not Protein Observed Fluorine NMR could be useful as a ligand discovery tool in a fragment screening capacity. Screening techniques using structure-based methods ideally address the three following questions in varying degree: 1) Does a particular compound bind to the protein target? 2) How well does it bind? And 3) Where on the protein does it bind? This manuscript sought to establish that PrOF NMR could be sufficiently used to address all three of these questions. Further, we sought to evaluate the efficacy of using PrOF NMR as a method for assessing the druggability of different binding sites by comparing hit rates for each site to other established screening tools. KIX, the model protein system that we used in this study, features two native transcription factor binding sites, which are both biologically relevant targets of interest.

2.1 **Introduction**

Protein-protein interactions (PPIs) are essential nodes in almost all biological processes within the cell. Modulation of these biomolecular interactions via small molecules is now a validated approach for therapeutic intervention[167-168] and chemical probe development to improve our understanding of biological systems.[169] Despite this validation, disruption of these interactions is difficult due to the large and plastic interfaces at the protein surface and lack of deep hydrophobic binding sites for accommodating traditional drug-like molecules.[16] This is particularly the case for the

transient and dynamic interactions found for many transcription factor-protein binding events that occur with modest affinity (K_d = 0.1-100 μM).

We recently reported on a protein-observed fluorine NMR method (PrOF NMR) for characterizing ligand binding at PPI sites.[41, 98] We employed singly-labeled fluorinated aromatic amino acids due to the enrichment of aromatic amino acids at PPI interfaces.[23, 170] In many cases, the modest protein structural perturbations from a single aryl hydrogen to fluorine substitution, the fast data acquisition for small to medium sized proteins, and the orthogonality of fluorine with biological nuclei, lead to a simplified and readily obtainable 1D-NMR spectrum reporting directly on protein structure.[74, 171] We used this method to characterize small molecule binding to three different epigenetic effector domains: bromodomains Brd4, BrdT, and BPTF.[98] We also studied the protein interaction domain of CBP/p300, KIX, by PrOF NMR and identified small molecule ligands 1-10 and 1G7.[41] Beyond model studies, PrOF NMR has yet to be tested in a full discovery format. Herein, we report the first full small molecule screen using PrOF NMR, evaluate its effectiveness for assessing protein druggability at two different binding sites within the same protein, KIX, and report structure activity relationships (SAR) from several newly discovered small molecules.

Molecules of low molecular weight and complexity, called fragments, have helped increase successful screening outcomes against difficult protein targets.[147] NMR has become a preferred method for screening for identifying low affinity ligands, in many cases using 1D-NMR ligand-based methods.[172-174] Analysis of fragment screen hit rates provides valuable information into the ligandability or "druggability" of the proteins,[175] and further development of fragments has led to molecules with more efficacious binding affinity and improved physicochemical properties than molecules obtained from traditional high-throughput screening.[176] To test the feasibility of using PrOF NMR in a fragment screening format, we chose the KIX domain.

KIX is a useful model protein for characterizing protein-small molecule interactions and studying the conformational plasticity of this dynamic protein domain. KIX binds to over a dozen transcription factors (e.g., MLL, E2A-PBX1, and CREB)

through two binding sites.[40, 177-178] These sites as well as new cryptic sites cannot readily be distiguished in a direct binding 1D-NMR ligand-based experiment. Prior screening efforts offer valuable ligand information for comparison with our PrOF NMR experiments. Two natural product depsides, sekikaic acid and lobaric acid, are reported as the most potent KIX inhibitors at the MLL-KIX PPI site (IC_{50} = 34 and 17 μM respectively). These compounds isolated from natural products were discovered after a screen of 50,000 commercial library compounds had resulted in no hits.[42] The only HSQC NMR screen resulted in two lead compounds, KG-122 (pamoic acid) and KG-501 from a 762-member peptidomimetic library (Fig. 1).[36, 179] Isoxazolidine derivatives, discovered first as general transcriptional activation domain mimics,[180] were later shown to also be KIX ligands.[181] Despite the paucity of potent ligands, both CREB-KIX, and E2A-KIX interactions have been proposed as important PPI targets for treating blood cancers due to their regulatory role in blood cell proliferation further motivating our research.[35, 177]

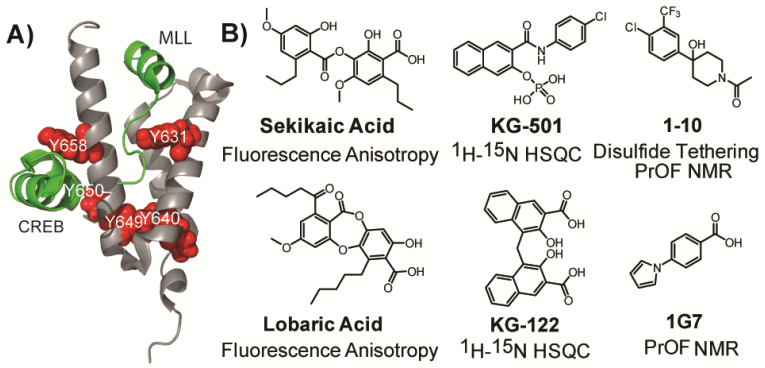


Figure 2.1 Solution structure of KIX and known ligands.
A) NMR solution structure of KIX (PDB: 2LXT) with natural peptides MLL and pKID (CREB). Tyrosine residues are shown as spheres. B) Structures of KIX ligands and their corresponding methods of discovery.

PPIs involving the KIX domain occur through two distinct and allosterically-coupled binding sites on this 3-helical bundle protein. These sites are often referred to as the MLL and CREB/c-Myb sites.[42] To conduct the NMR screen, we replaced native tyrosine residues of KIX with 3-fluorotyrosine (3FY). The MLL site contains one

tyrosine (Y631), which makes a critical contact with the native protein. A sequence analysis by Näär et al. of KIX domains shows this residue is conserved in humans, *C. elegans* and *drosophila*.[182] Y649 is evolutionarily conserved in all KIX domains. Fluorine resonances from Y649, Y650, and Y658 experience significant chemical shifts when CREB binds to KIX as well as Y631 from allostery.[41] Consistent with natural ligands, our computed druggability analysis using SiteMap identified two druggable sites near the tyrosine residues at both the MLL (10.8 Å from Y631) and CREB/Myb sites (4-10 Å from Y649, Y650, and Y658).[183] Y640, which structurally stabilizes KIX via a cation-π interaction with R600,[184] is found in 97% of KIX domains, and may represent a new small molecule site for regulating protein conformation.[182] Importantly, singly fluorinated aromatic residues in KIX only modestly perturb secondary structure and ligand binding.[41]

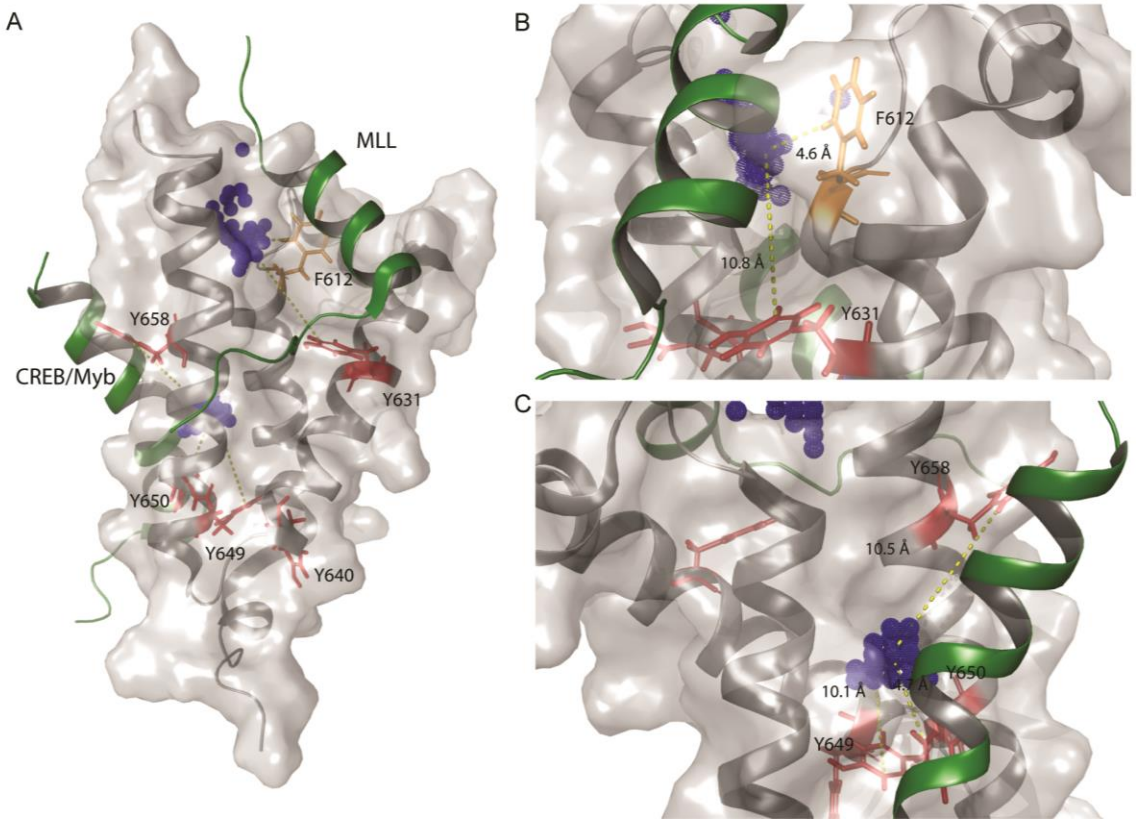


Figure 2.2 SiteMap analysis of KIX (PDB 2AGH).
Tyrosine and phenylalanine are shown as sticks. Binding sites identified by SiteMap are represented as spheres. Native peptides MLL and Myb are shown as ribbons. A) Full protein view portraying both native binding sites identified by SiteMap. B) MLL binding

site showing both F612 and Y631 in close proximity to the binding site. C) CREB/Myb binding site showing Y649, Y650, and Y658 in close proximity to the identified binding site.

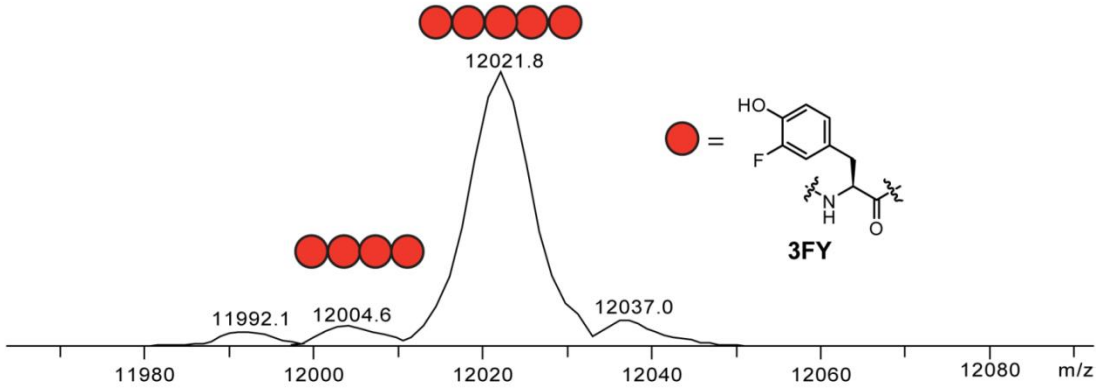


Figure 2.3. The deconvoluted ESI-mass spectrogram of a 3FY KIX sample demonstrating high fluorine incorporation.
The peak at 12004.6 is an 80% labeled protein, the peak at 12021.8 is the 100% labeled protein containing all five 3FY residues, and the peak at 12037.0 is consistent with a small percentage of an oxidized product that was presumed to occur from methionine oxidation. For high labeling, 80 mg 3FY per liter is added to the defined media.

2.2 **Results and Discussion**

We expressed 3FY-labeled KIX (12 kDa) in good yield (70 mg/L) with high labeling efficiency (>98%) for the screen.[183] Our fragment library was generated from the Maybridge rule of 3 commercial set, combining compounds into 85 mixtures of five or six compounds at a total stock concentration of 33.3 mM per compound in DMSO. As a positive control, we tested a known ligand, KG-501 (833 µM), and identified a binding interaction, both in isolation, consistent with prior results,[41] and in a mixture based on chemical shift perturbation of Y631 (Fig. 2B). For the NMR screen we used 40 µM KIX (~20 mg protein total). Chemical shift information was acquired in five minutes yielding approximately 510 minutes of experiment time for screening the 85 mixtures, including additional short reference experiments for each mixture. This experiment time is faster than ^{1}H-^{15}N SOFAST HMQC NMR experiments for similar sized proteins.[185] All mixtures were screened at 833 µM small molecule and 2.5% DMSO. Statistical cut-offs for chemical shift perturbation were set to two standard deviations from the average perturbations from the screen yielding 15 mixtures (Fig. 2C). Each mixture was

subsequently deconvoluted by individual analysis of each compound leading to four verified ligands (**1-4**, 0.8% hit rate). The reduction in hits was due to apparent additive effects of fragments which prevented the identification of a sole compound responsible for chemical shift perturbations in a given mixture.

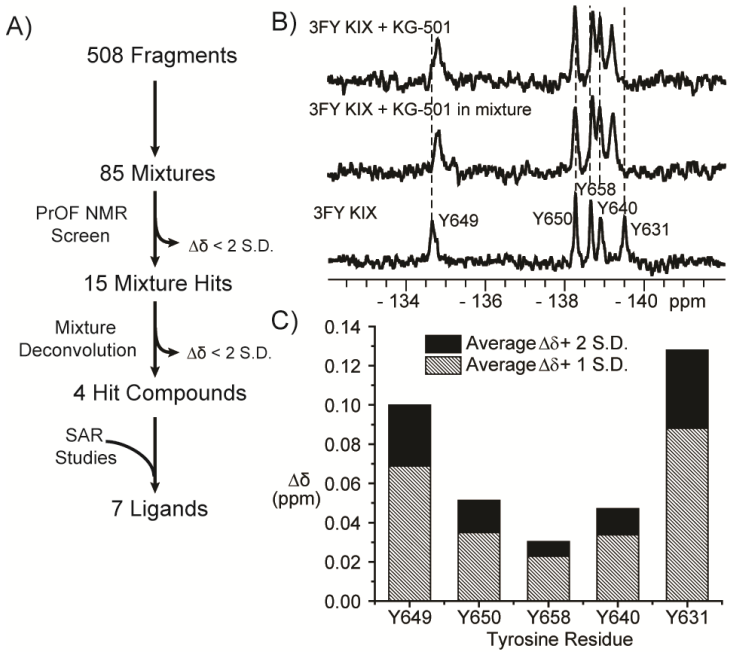


Figure 2.4 Fragment screening by PrOF NMR.
A) Flowchart for screening of fragment library for KIX ligands. 508 fragments were screened using PrOF NMR, yielding 15 mixtures hits and seven final ligand after deconvolution and SAR studies. B) KG-501 was used to test for KIX ligand identification in isolation or a mixture of six total compounds. C) Statistical cut-offs for chemical shift perturbations found in the screen. Samples with chemical shift pertubations greater than two standard deviations were identified as hits.

Ligand titration experiments via PrOF NMR were performed to determine the dissociation constants for the small molecule obtained by monitoring changes in chemical shift (Δδ). Three of the four ligands (**2**, **3**, and **4**) discovered from the screen were found to have low mM binding affinities for KIX. These compounds contained either an aryl or phenylacetic acid group (Table 1). Small molecules **3** and **4** exhibited a binding isotherm consistent with one-to-one binding while **2** potentially exhibited higher-order binding above 2 mM based on the binding isotherm generated. Consequently the K_d for **2** was

estimated based on fitting the data up to 2 mM. We ruled out activity due to small molecule aggregation above 2 mM based on well resolved small molecule resonances in the presence and the absence of detergent.[183]

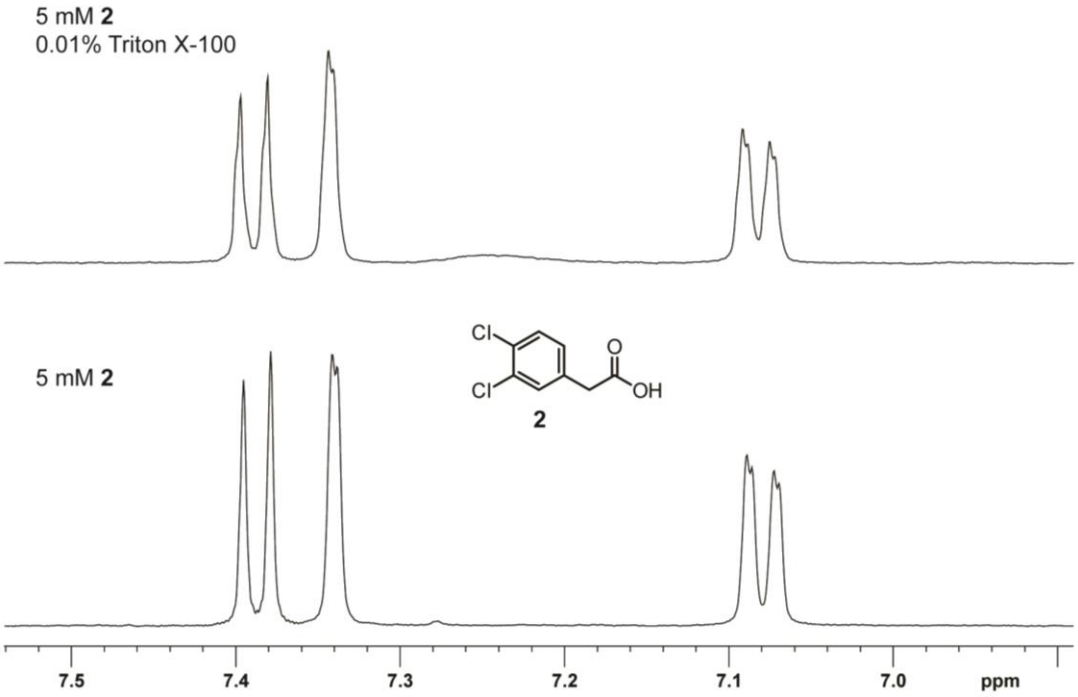


Figure 2.5 Aromatic region of 2 in the presence and absence of detergent in aqueous buffer.
Buffer contains 50 mM HEPES and 100 mM NaCl at pH 7.2. Linewidths are not significantly perturbed in the presence of detergent, suggesting a lack of small molecule aggregation.

Given that KG-122 is a diacid compound that binds to KIX in the MLL site, it is possible that multiple copies of the monoacid **2** could bind in the same region on the protein. Y631, which is featured in the MLL site, was the most sensitive reporter for binding and was therefore used to calculate the K_d for **2, 3,** and **4**. In contrast, Y649, Y650, and Y658, which are featured in the CREB binding site, were significantly less responsive to ligand binding with the exception of **1** which perturbed resonances for Y649, Y650, and Y631, potentially indicating binding at the CREB site.

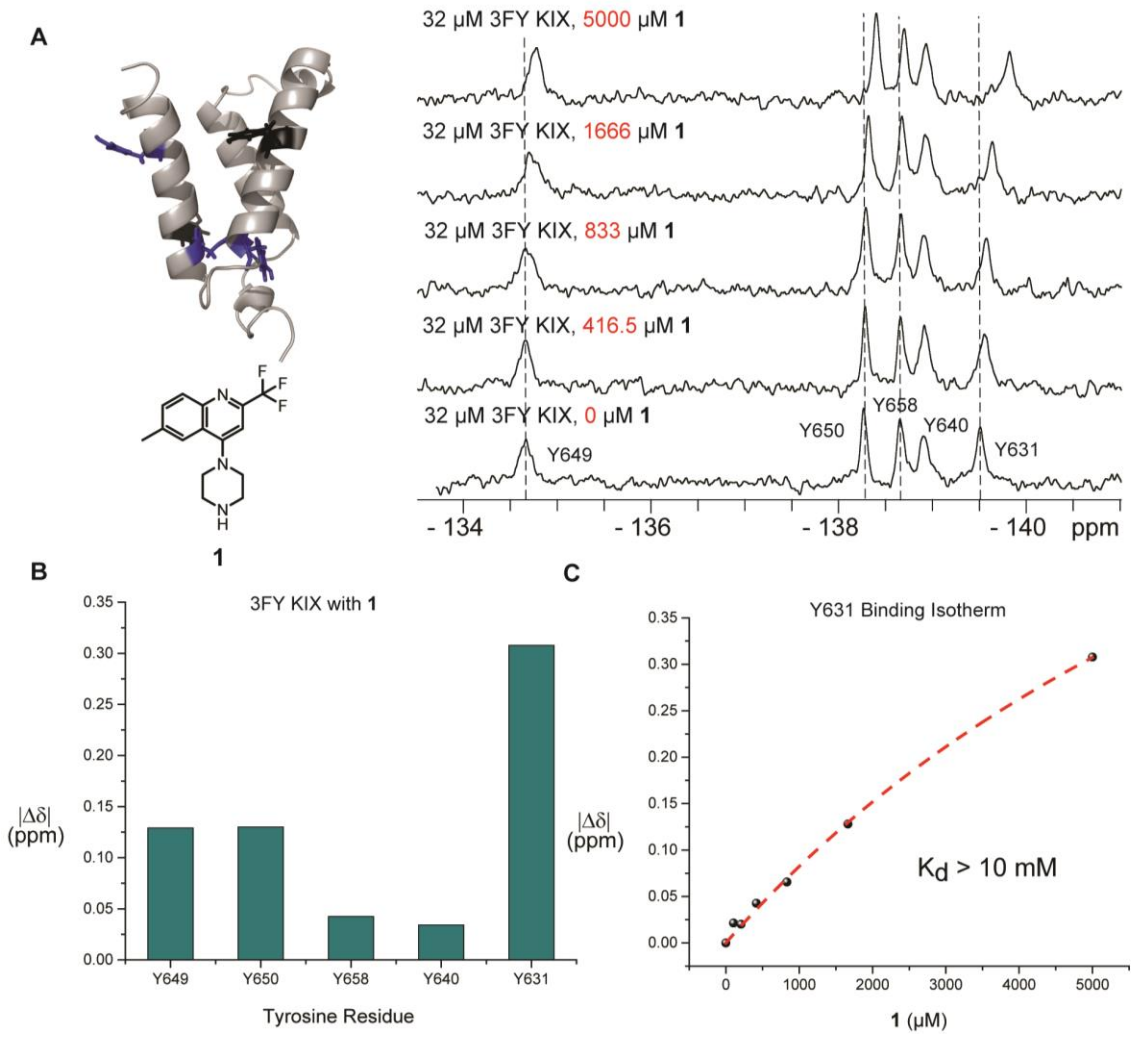


Figure 2.6 ^{19}F NMR spectral analysis of 3FY KIX in the presence of increasing concentrations of 1.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. Tyrosine residues shown as sticks experienced statistically significant perturbations at low concentrations (833 μM) but failed to approach saturation, were not significantly perturbed, or were significantly perturbed and approached saturation. B) Absolute value of chemical shift perturbations for all 3FY KIX tyrosine resonances at 5 mM **1**. C) Binding isotherm of Y631 perturbation for the titration of **1**.

Table 2.1. 3FY-KIX 19F NMR chemical shift perturbations at varying **1** concentrations

[**1**] (μM)	Chemical shift (ppm)	Y649	Y650	Y658	Y640	Y631
0	Chemical shift (ppm)	-134.6690	-138.2668	-138.6544	-138.9044	-139.5114
104	Δδ (ppm)	0.0212	0.0076	0.0090	0.0091	0.0214
208	Δδ (ppm)	-0.0076	0.0182	0.0055	-0.0080	0.0201
416.5	Δδ (ppm)	-0.0060	0.0161	0.0060	0.0092	0.0427
833	Δδ (ppm)	-0.0032	0.0238	0.0135	0.0044	0.0656
1666	Δδ (ppm)	0.0321	0.0523	0.0213	0.0160	0.1280
5000	Δδ (ppm)	0.1291	0.1300	0.0424	0.0342	0.3078

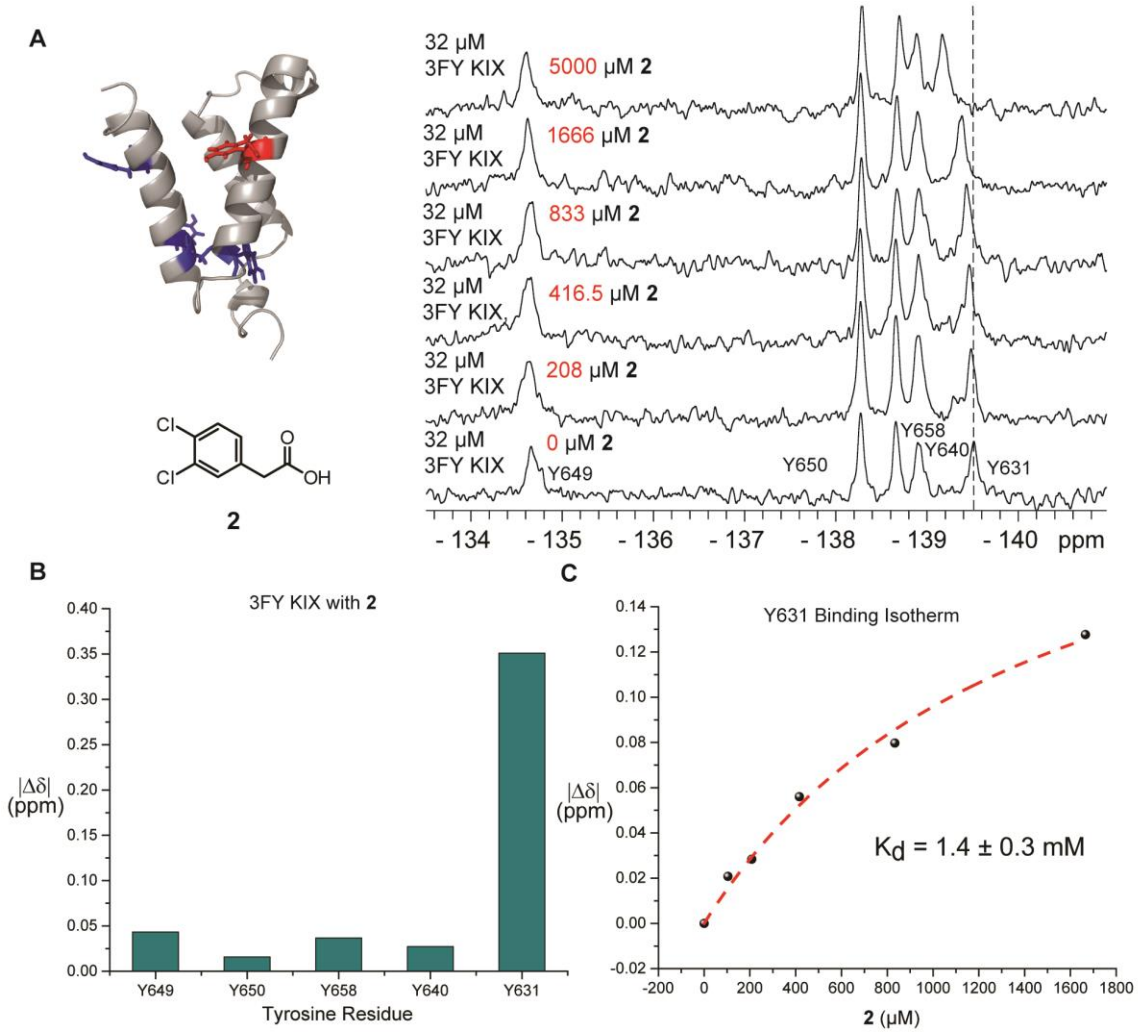


Figure 2.7. ^{19}F NMR spectral analysis of 3FY KIX in the presence of increasing concentrations of 2.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. Tyrosine residues shown as sticks experienced statistically significant perturbations at low concentrations (833 μM) but failed to approach saturation, were not significantly perturbed, or were significantly perturbed and approached saturation. B) Absolute value of chemical shift perturbations for all 3FY KIX tyrosine resonances at 5 mM **2**. C) Binding isotherm of Y631 perturbation for the titration of **2**.

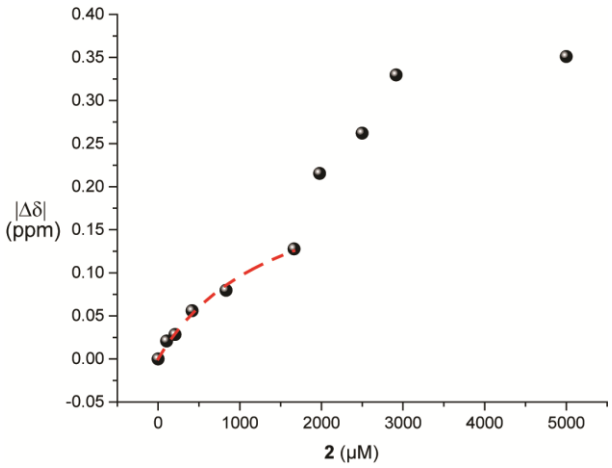


Figure 2.8. Full titration of **2**, demonstrating non-first order binding behavior above 2 mM.

Table 2.2. 3FY-KIX 19F NMR chemical shift perturbations at varying **2** concentrations

[**2**] (μM)	Chemical shift (ppm)	Y649	Y650	Y658	Y640	Y631
0	Chemical shift (ppm)	-134.6690	-138.2668	-138.6544	-138.9044	-139.5114
104	Δδ (ppm)	-0.0066	0.0054	0.0030	-0.0161	-0.0208
208	Δδ (ppm)	-0.0259	0.0051	0.0052	0.0037	-0.0284
416.5	Δδ (ppm)	-0.0962	-0.0077	-0.0057	-0.0040	-0.0560
833	Δδ (ppm)	0.0051	0.0137	0.0156	0.0046	-0.0797
1666	Δδ (ppm)	-0.0477	0.0053	0.0163	-0.0090	-0.1277
1978	Δδ (ppm)	-0.0476	0.0065	0.0196	-0.0134	-0.2154
2499	Δδ (ppm)	-0.0333	0.0158	0.0341	-0.0169	-0.2621
2916	Δδ (ppm)	-0.0523	0.0158	0.0391	-0.0319	-0.3297
5000	Δδ (ppm)	-0.0434	0.0158	0.0368	-0.0272	-0.3510

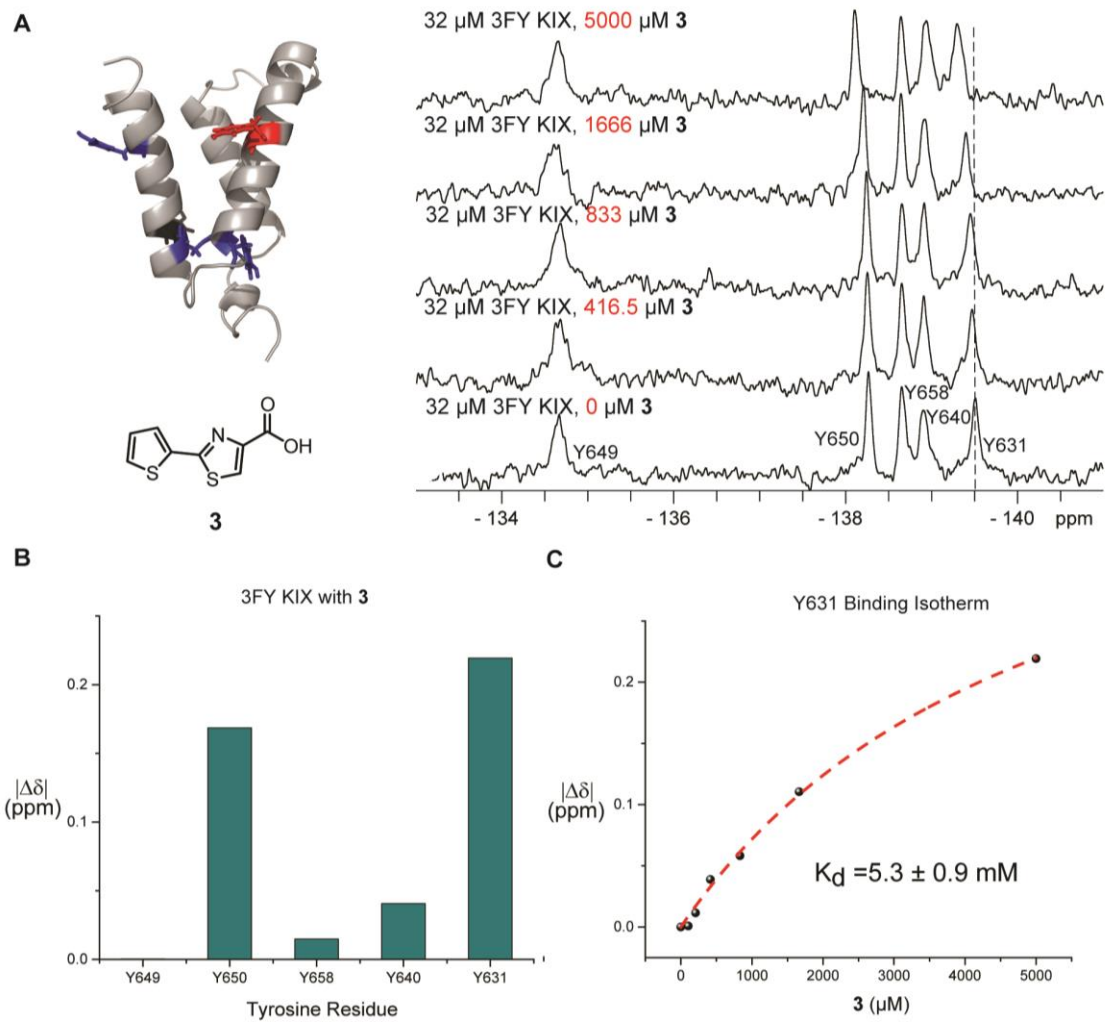


Figure 2.9. ^{19}F NMR spectral analysis of 3FY KIX in the presence of increasing concentrations of 3.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. Tyrosine residues shown as sticks experienced statistically significant perturbations at low concentrations (833 μM) but failed to approach saturation, were not significantly perturbed, or were significantly perturbed and approached saturation. B) Absolute value of chemical shift perturbations for all 3FY KIX tyrosine resonances at 5 mM **3**. C) Binding isotherm of Y631 perturbation for the titration of **3**.

Table 2.3. 3FY-KIX 19F NMR chemical shift perturbations at varying **3** concentrations

Table S4. 3FY-KIX ^{19}F NMR chemical shift perturbations at varying **3** concentrations.						
[**3**] (μM)	Chemical shift (ppm)	Y649	Y650	Y658	Y640	Y631
0	Chemical shift (ppm)	-134.6690	-138.2668	-138.6544	-138.9044	-139.5114
104	Δδ (ppm)	-0.0183	0.0000	0.0056	0.0185	-0.0008
208	Δδ (ppm)	-0.0306	0.0015	0.0062	0.0164	-0.0115
416.5	Δδ (ppm)	0.0142	-0.0110	0.0001	0.0067	-0.0388
833	Δδ (ppm)	0.0158	-0.0226	0.0018	0.0112	-0.0582
1666	Δδ (ppm)	-0.0602	-0.0579	-0.0090	-0.2816	-0.1105
5000	Δδ (ppm)	-0.0003	-0.1686	-0.0149	0.0406	-0.2193

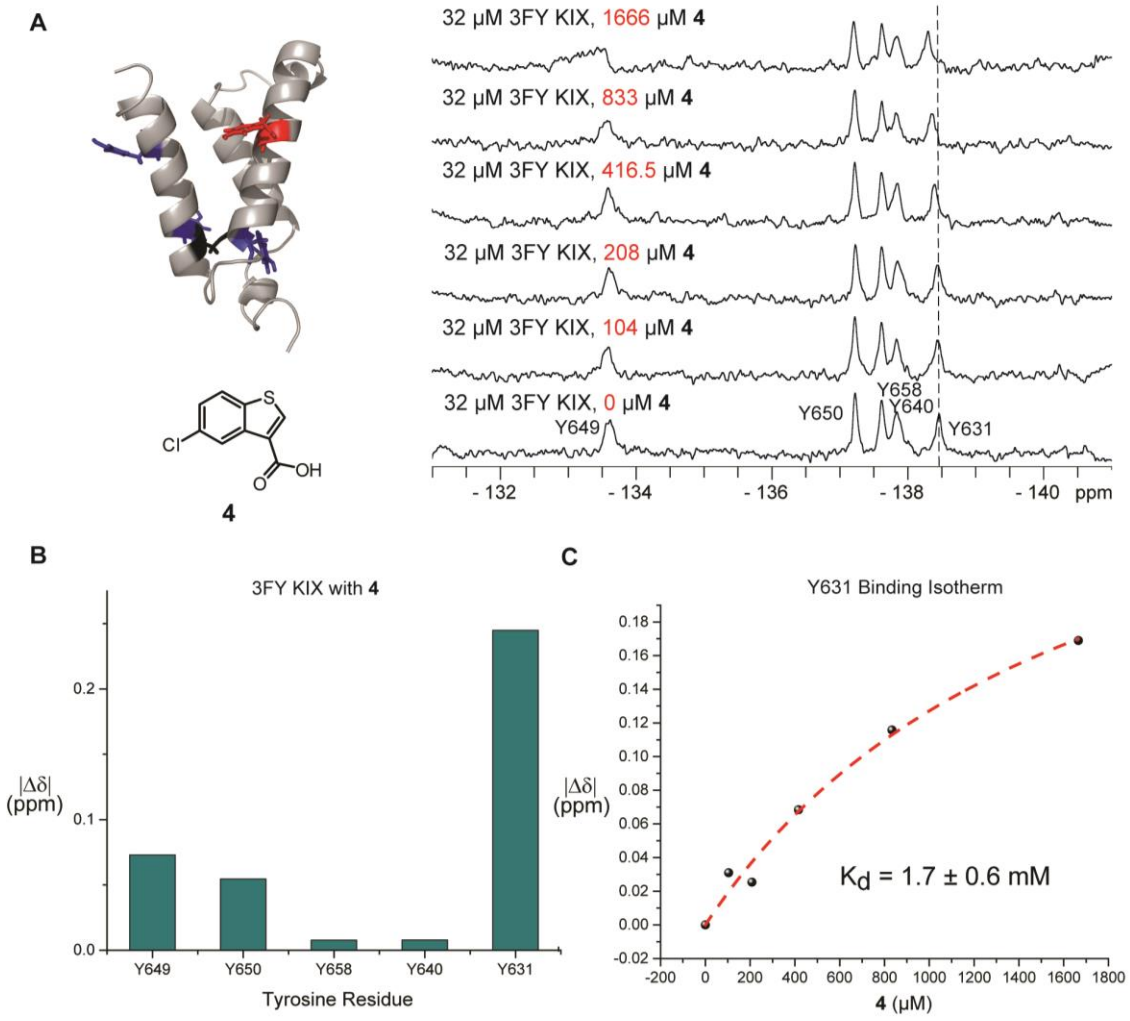


Figure 2.10. ^{19}F NMR spectral analysis of 3FY KIX in the presence of increasing concentrations of 4.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. Tyrosine residues shown as sticks experienced statistically significant perturbations at low concentrations (833 μM) but failed to approach saturation, were not significantly perturbed, or were significantly perturbed and approached saturation. B) Absolute value of chemical shift perturbations for all 3FY KIX tyrosine resonances at 5 mM **4**. C) Binding isotherm of Y631 perturbation for the titration of **4**.

Table 2.4. 3FY-KIX 19F NMR chemical shift perturbations at varying **4** concentrations.

[**4**] (μM)	Chemical shift (ppm)	Y649	Y650	Y658	Y640	Y631
0	Chemical shift (ppm)	-134.6653	-138.2711	-138.6618	-138.8869	-139.5090
104	Δδ (ppm)	-0.0255	-0.0038	-0.0037	-0.0062	-0.0310
208	Δδ (ppm)	-0.0225	0.0017	-0.0035	0.0061	-0.0254
416.5	Δδ (ppm)	-0.0329	-0.0056	-0.0049	0.0071	-0.0684
833	Δδ (ppm)	-0.0448	-0.0049	0.0006	-0.0194	-0.1158
1666	Δδ (ppm)	-0.0776	-0.0223	-0.0001	-0.0057	-0.1690

As a follow-up to the screen, the SAR from the four hits were examined to identify structural motifs important for binding. Four trifluoromethylquinoline derivatives of **1** were purchased. These compounds were of interest due to small but significant chemical shift perturbations from **1** of Y649, Y650, Y658, and Y631, which were similar to the chemical shift effects observed upon CREB binding to KIX suggesting a possible allosteric binding mode.[4a] One of the four compounds (KM07508) yielded comparable chemical shift perturbations as **1**.[183] However, as **1** displayed poor binding behavior in titration experiments (K_d > 10 mM) due to lack of saturated binding and in 2D NMR experimental results explained below, SAR studies with this ligand class were not pursued further in lieu of the investigation into aryl and phenylacetic acids.

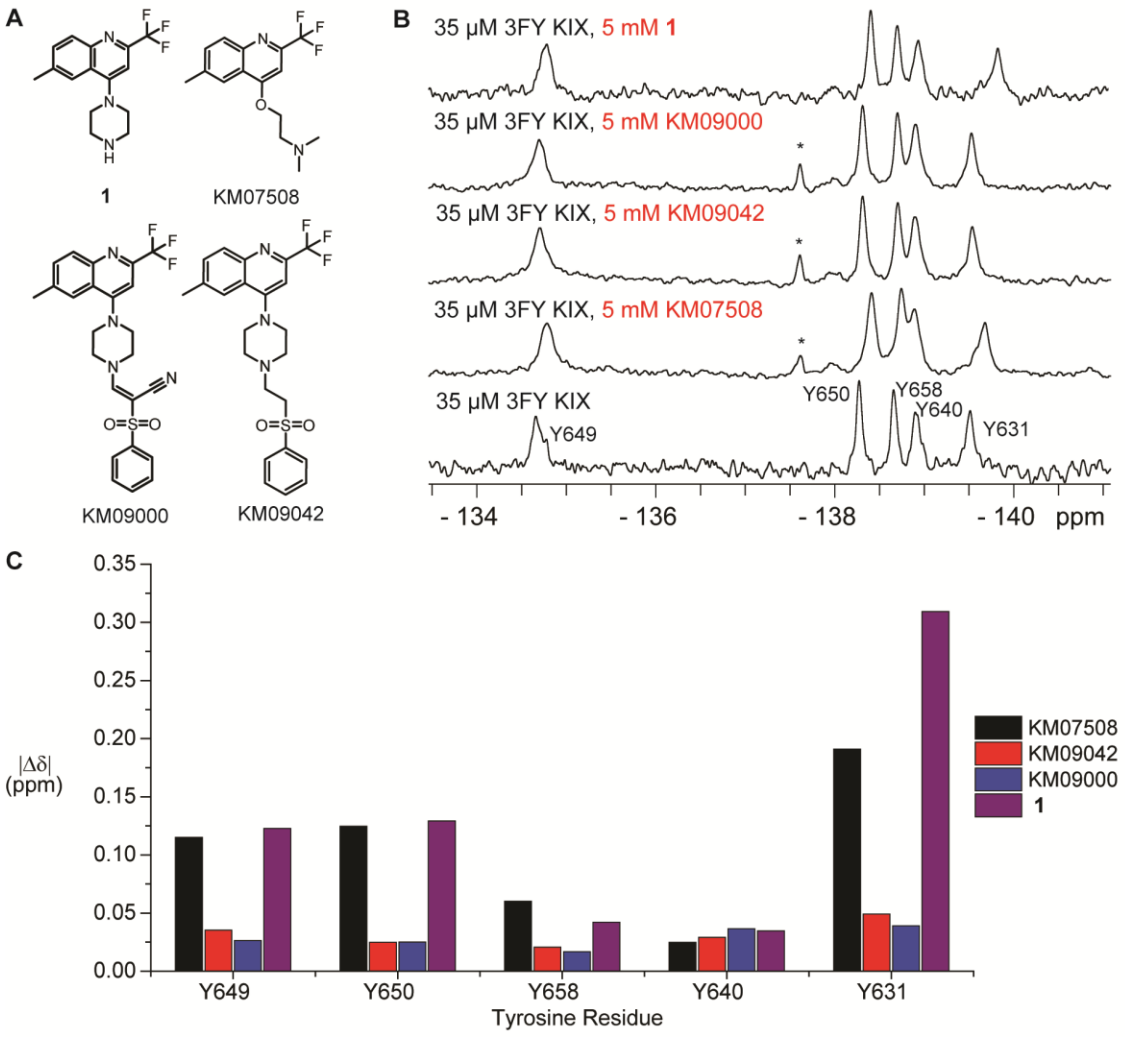


Figure 2.11 SAR analysis of 1 analogs.
A) Structures of **1** and corresponding analogs from Maybridge. B) Stacked PrOF NMR spectra with 5 mM ligands. Intensity scale for 3FY KIX and 1 spectra were increased to compensate for a larger number of scans for the other three experiments. * denotes a partial degradation peak of KIX. C) Absolute value of chemical shift perturbations of all tyrosine residues with 5 mM ligands.

The other three discovered ligands contained an aryl or phenylacetic acid motif as a potential pharmacaphore. The presence of a carboxylic acid has been found in other KIX ligands that bind to the MLL site (sekikaic acid, lobaric acid, KG-122, and 1G7, Fig 1). To evaluate specificity over other acidic molecules and to probe for false negatives, several compounds from library mixtures were retested. One of the four compounds

(9B11, **5**), a biaryl carboxylic acid, was anticipated to have been a false negative due to its similar structure to KIX ligand 1G7. The other three carboxylic acids were selected to test if these aryl and phenylacetic acids represented a class of ligands that bound KIX while other carboxylic acids would not (i.e., an aliphatic (**6**), a phenethyl (**7**), and a thiazole (**8**) carboxylic acid). Only **5** yielded significant changes in chemical shift, while the other three exhibited minimal to no apparent binding. Titration of **5** yielded a K_d of 1.6 ± 0.3 mM (Table 1). Low mM K_d values and ligand efficiencies from 0.19 to 0.32 kcal mol^{-1} non-hydrogen $atom^{-1}$ were calculated for each fragment. As a comparison, the reported median ligand efficiency was 0.32 for a sample set of 480 ligand-target pairs.[150] Our ligand efficiencies are equal to or slightly lower than this value, supporting the difficulty of this target site.[42]

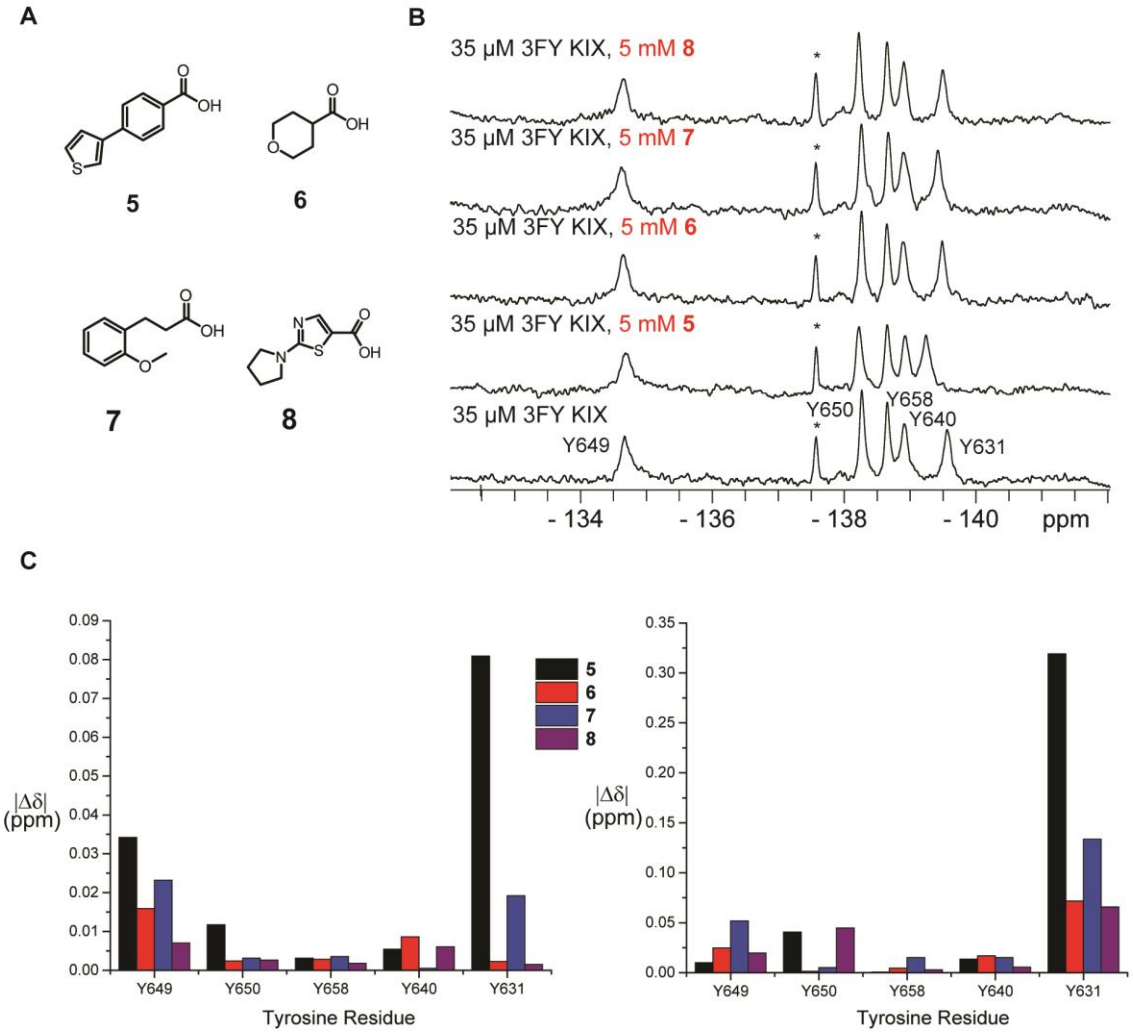


Figure 2.12 SAR analysis of carboxylic acids from screen.
A) Structures of four carboxylic acids being retested. B) Stacked PrOF NMR spectra with 5 mM ligands. * denotes a partial degradation peak from KIX. C) Absolute value of chemical shift perturbations of all tyrosine residues with 833 μM (left) and 5 mM ligands (right).

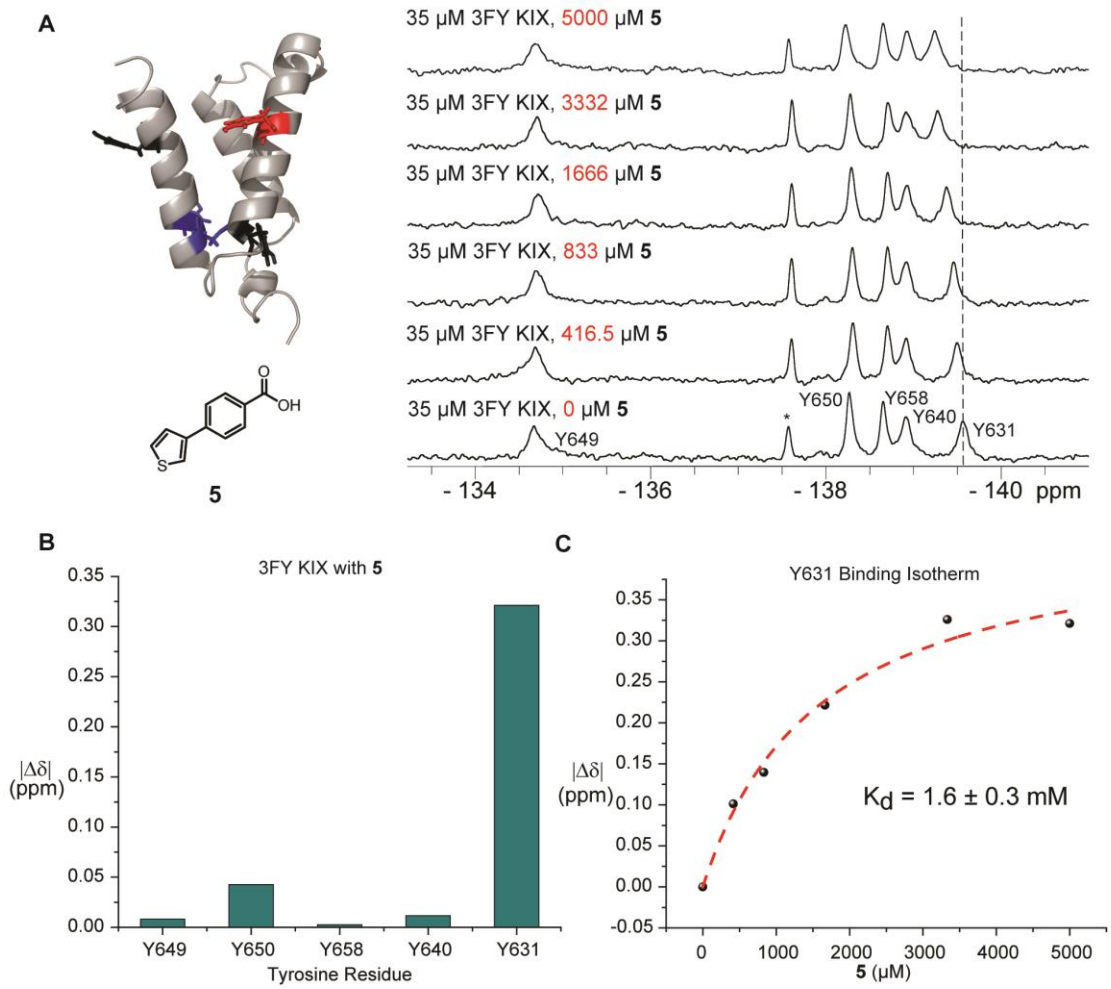


Figure 2.13 ^{19}F NMR spectral analysis of 3FY KIX in the presence of increasing concentrations of 5.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. Tyrosine residues shown as sticks experienced statistically significant perturbations at low concentrations (833 μM) but failed to approach saturation, were not significantly perturbed, or were significantly perturbed and approached saturation. B) Absolute value of chemical shift perturbations for all 3FY KIX tyrosine resonances at 5 mM **5**. C) Binding isotherm of Y631 perturbation for the titration of **5**.

Table 2.5. 3FY-KIX 19F NMR chemical shift perturbations at varying **5** concentrations

Table S6: 3FY-KIX ^{19}F NMR chemical shift perturbations at varying **5** concentrations.						
[**5**] (μM)	Chemical shift (ppm)	Y649	Y650	Y658	Y640	Y631
0	Chemical shift (ppm)	-134.6704	-138.2670	-138.6540	-138.9117	-139.5613
416.5	Δδ (ppm)	-0.0209	0.0068	0.0145	-0.0310	-0.1013
833	Δδ (ppm)	-0.0173	0.0006	0.0153	-0.0296	-0.1396
1666	Δδ (ppm)	0.0111	-0.0055	0.0155	-0.0198	-0.2215
3332	Δδ (ppm)	-0.0029	-0.0275	0.0138	-0.0374	-0.3259
5000	Δδ (ppm)	0.0081	-0.0427	-0.0026	0.0116	-0.3211

To expand on the investigation of aryl + acid SAR, nine arylacetic and benzoic acid derivatives were tested as analogs of molecule **2** to inform optimization studies, based on the affinity of **2** and the readily modifiable structure of phenylacetic acids. Of the nine compounds, only the biphenyl (PAA5) yielded a pronounced chemical shift perturbation close to one standard deviation above the average at 5 mM,[183] despite the structural similarity of these analogues to **2**, e.g., 3-chloro-4-hydroxy phenylacetic acid (PAA2), phenylacetic acid (PAA3), and 4-hydroxyphenylacetic acid (PAA4) (Fig. 3). Each analog was also analyzed at 2 mM and 833 μM, due to the potential for higher order binding at high concentrations (> 2 mM). The trend remained consistent with all compounds aside from PAA5 yielding insignificant changes in chemical shift. The K_d for PAA5 binding to KIX was 6.9 ± 0.7 mM. These results suggest that hydrophobic groups at the para position on the phenylacetic acid play a key role in ligand binding.

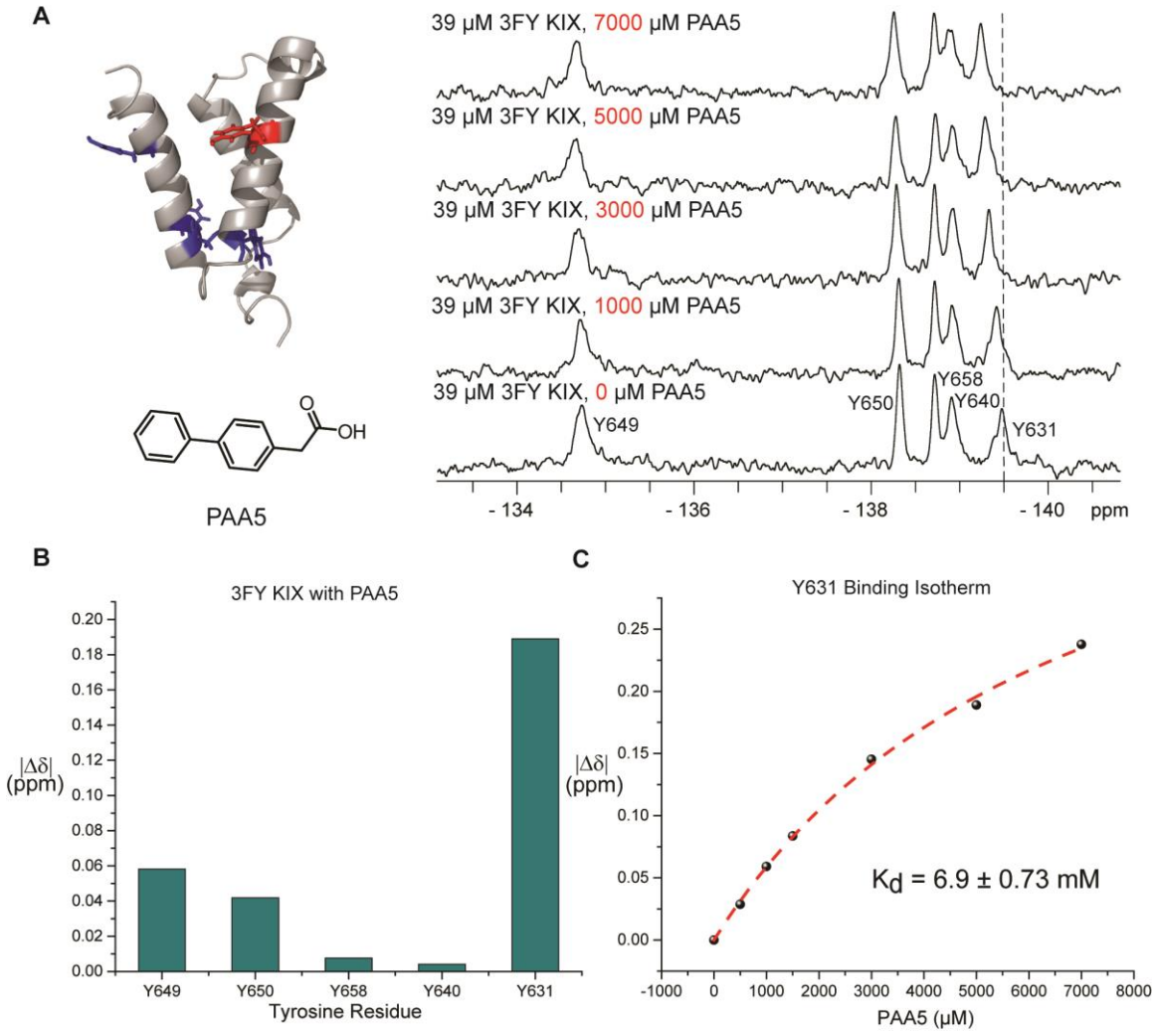


Figure 2.14. ^{19}F NMR spectral analysis of 3FY KIX in the presence of increasing concentrations of PAA5.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. Tyrosine residues shown as sticks experienced statistically significant perturbations at low concentrations (833 μM) but failed to approach saturation, were not significantly perturbed, or were significantly perturbed and approached saturation. B) Absolute value of chemical shift perturbations for all 3FY KIX tyrosine resonances at 5 mM PAA5. C) Binding isotherm of Y631 perturbation for the titration of PAA5.

Table 2.6. 3FY-KIX ^{19}F NMR chemical shift perturbations at varying PAA5 concentrations

[PAA5] (μM)	Chemical shift (ppm)	Y649	Y650	Y658	Y640	Y631
0	Chemical shift (ppm)	-134.7310	-138.3225	-138.7184	-138.9099	-139.4771
500	Δδ (ppm)	-0.0137	-0.0097	-0.0057	0.0021	-0.0287
1000	Δδ (ppm)	-0.0183	-0.0084	0.0001	0.0019	-0.0591
1500	Δδ (ppm)	0.0097	-0.0152	-0.0039	0.0105	-0.0837
3000	Δδ (ppm)	-0.0290	-0.0360	-0.0015	0.0177	-0.1453
5000	Δδ (ppm)	-0.0582	-0.0418	0.0076	0.0042	-0.1890
7000	Δδ (ppm)	-0.0591	-0.0645	-0.0067	-0.0363	-0.2378

Phenylacetic acids are a common scaffold found in several nonsteroidal anti-inflammatory drugs. As an additional analysis of SAR, naproxen, ibuprofen, and flurbiprofen were used to further support the model that para-substituted phenylacetic acid derivatives would bind to KIX. All three compounds were analyzed by PrOF NMR. The chemical shift perturbations from enantiopure naproxen were under the statistical cutoff at low concentrations. Racemic ibuprofen yielded significant chemical shift perturbations (Δδ), while racemic flurbiprofen yielded the largest Δδ for Y631 observed thus far (0.38 ppm at 5 mM, K_d = 5.9 ± 0.6 mM, Fig. 3). Subsitution at the α-position affords an additional site of diversity and both stereochemistry and elaboration of side chains can be investigated as starting points for structure activity studies.

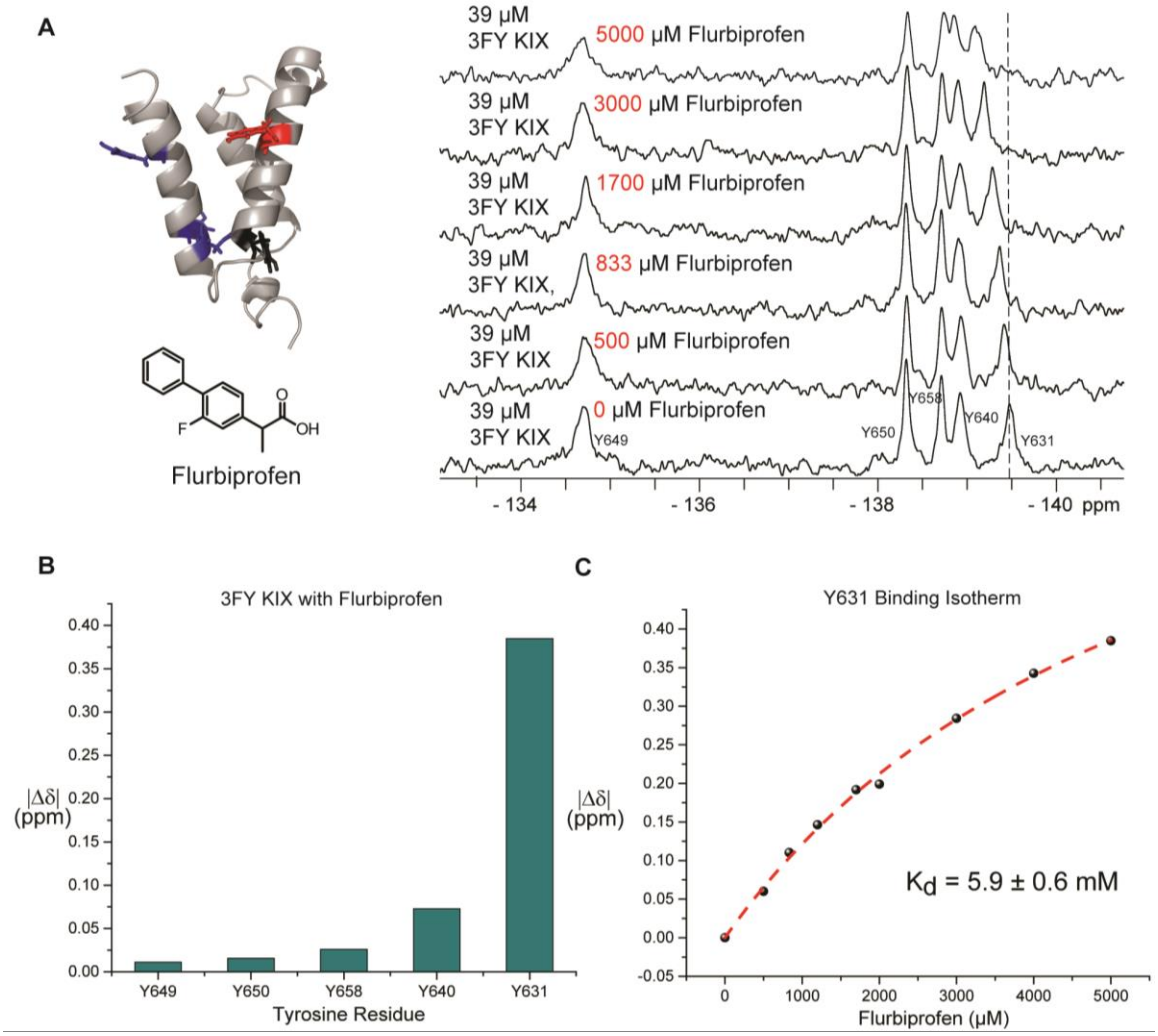


Figure 2.15. ^{19}F NMR spectral analysis of 3FY KIX in the presence of increasing concentrations of flurbiprofen.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. Tyrosine residues shown as sticks experienced statistically significant perturbations at low concentrations (833 μM) but failed to approach saturation, were not significantly perturbed, or were significantly perturbed and approached saturation. B) Absolute value of chemical shift perturbations for all 3FY KIX tyrosine resonances at 5 mM flurbiprofen. C) Binding isotherm of Y631 perturbation for the titration of flurbiprofen.

Table 2.7. 3FY-KIX ^{19}F NMR chemical shift perturbations at varying flurbiprofen concentrations

[Flu] (μM)	Chemical shift (ppm)	Y649	Y650	Y658	Y640	Y631
0	Chemical shift (ppm)	-134.6964	-138.3180	-138.7125	-138.9233	-139.4756
500	Δδ (ppm)	0.0009	0.0062	0.0020	0.0059	-0.0601
833	Δδ (ppm)	0.0261	-0.0035	-0.0012	-0.0205	-0.1104
1200	Δδ (ppm)	0.0.0229	0.0049	0.0057	0.0019	-0.1464
1700	Δδ (ppm)	0.0298	0.0034	0.0023	-0.0045	-0.1918
2000	Δδ (ppm)	0.0323	0.0093	0.0084	-0.0165	-0.1991
3000	Δδ (ppm)	-0.0154	0.0120	0.0061	-0.0231	-0.2842
4000	Δδ (ppm)	-0.0320	0.0050	0.0146	-0.0478	-0.3427
5000	Δδ (ppm)	0.0112	0.0156	0.0259	-0.0729	-0.3848

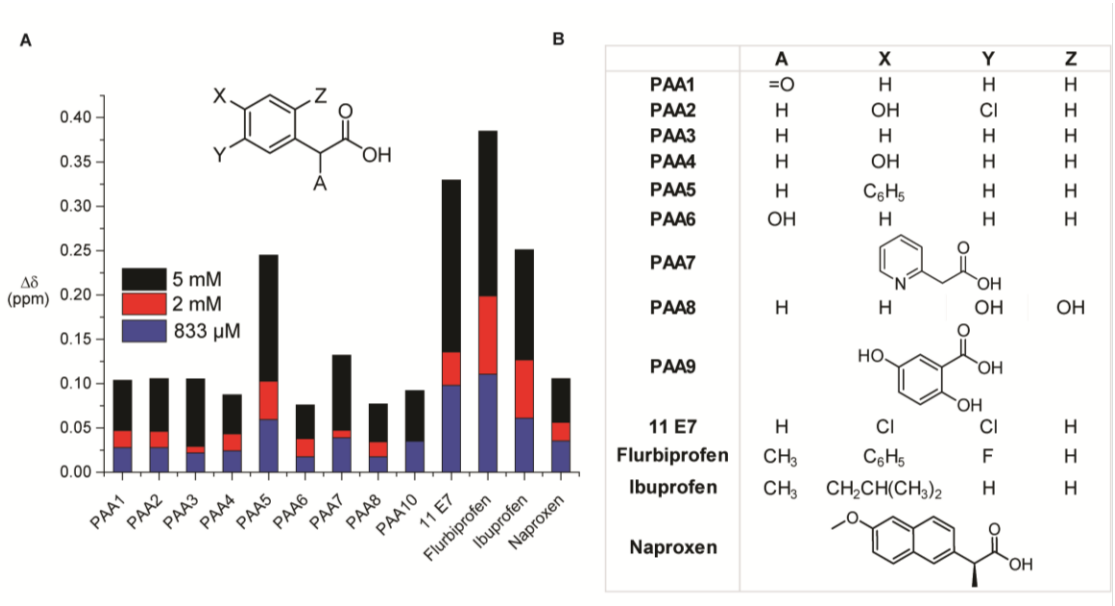


Figure 2.16 SAR analysis of phenylacetic/arylacetic acid derivative compounds.
A) Chemical shift perturbations at three different concentrations of small molecule. B) Interpretation table for phenylacetic acid derivatives. Structures of compounds that did not match the general template are drawn.

As a complementary method to evaluate the ligand binding site on KIX and rule out concerns of generating false positives from fluorine incorporation, we studied the binding of the original four ligands (**1-4**) by ^{1}H-^{15}N HSQC NMR using uniformly ^{15}N-labeled KIX. Experiments were performed at 0, 1, and 5 mM ligand concentration for **1**, **2**, and **3** while **4** was characterized at 0, 0.5, and 1 mM due to low solubility above 2 mM. Residues whose chemical shifts were perturbed by one and two standard deviations above the average perturbation in a residue were mapped onto the protein structure to infer the corresponding binding site of each ligand.

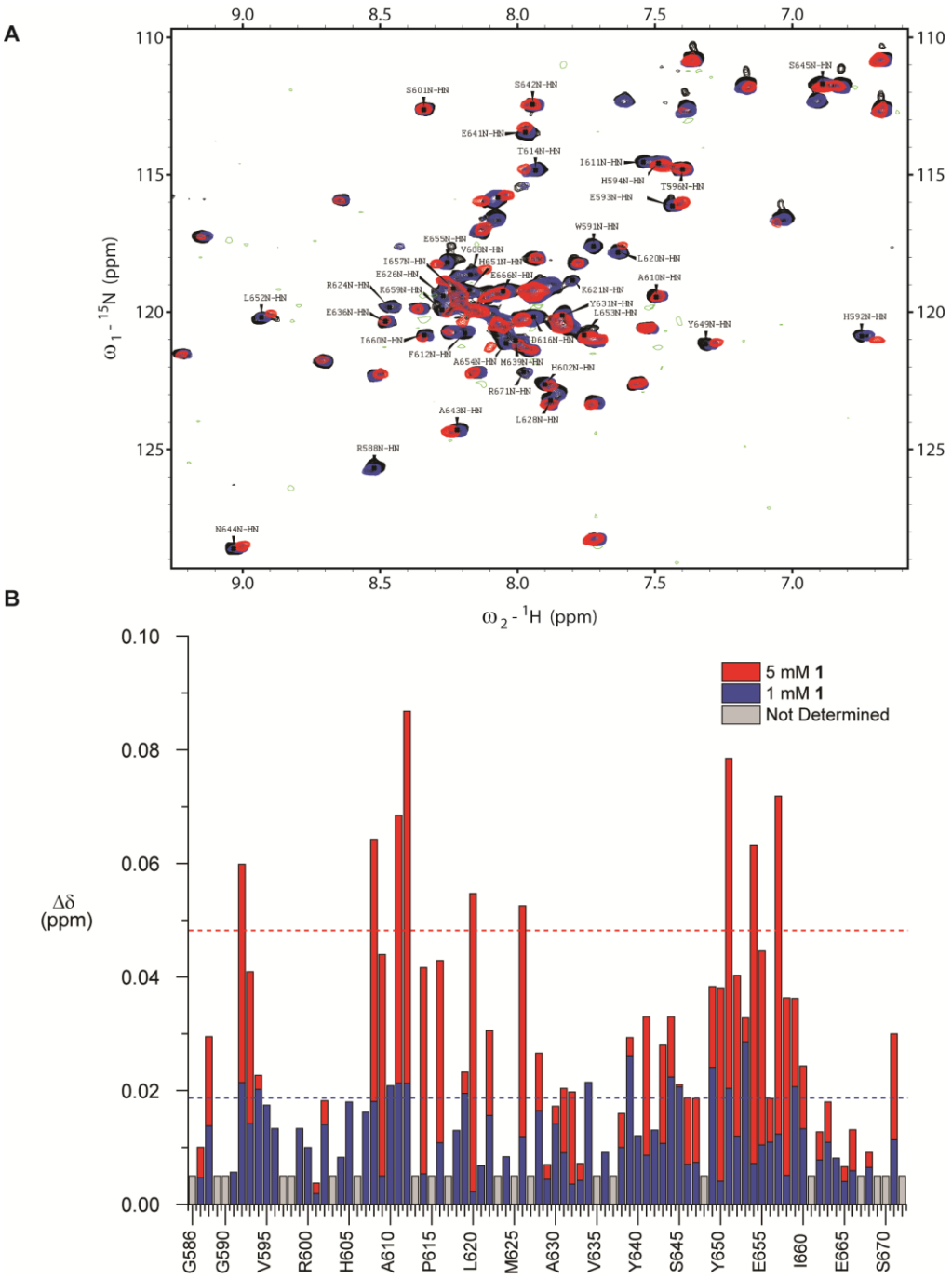


Figure 2.17 ^{1}H-^{15}N HSQC of 1 binding to KIX.
A) ^{15}N KIX alone, ^{15}N KIX + 1 mM **1**, and ^{15}N KIX + 5 mM **1** are shown in the spectrum. The peak arrows refer to KIX without ligand. B) Chemical shift perturbation mapping of KIX residues with 1 mM and 5 mM **1**. Residues that were not detected or not assigned are designated as ND. Dashed lines represent the average chemical shift perturbation + one standard deviation at the corresponding concentration. Assignments previously obtained by Majmudar et al.[42]

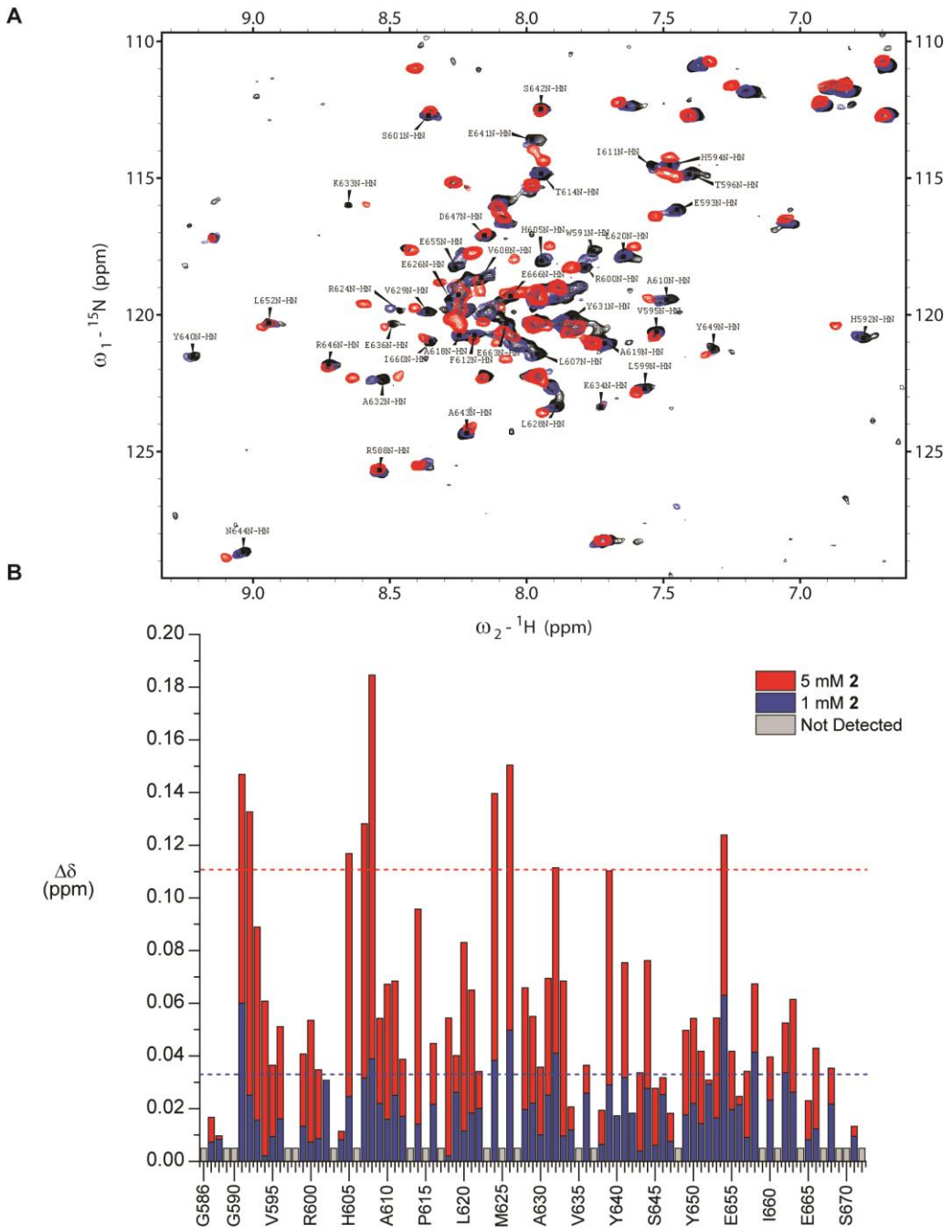


Figure 2.18. ^{1}H-^{15}N HSQC of 2 binding to KIX.
A) ^{15}N KIX alone, ^{15}N KIX + 1 mM **2**, and ^{15}N KIX + 5 mM **2** are shown in the spectrum. The peak arrows refer to KIX without ligand. B) Chemical shift perturbation mapping of KIX residues with 1 mM and 5 mM **2**. Residues that were not detected or not assigned are designated as ND. Dashed lines represent the average chemical shift perturbation + one standard deviation at the corresponding concentration. Assignments previously obtained by Majmudar et al.[42]

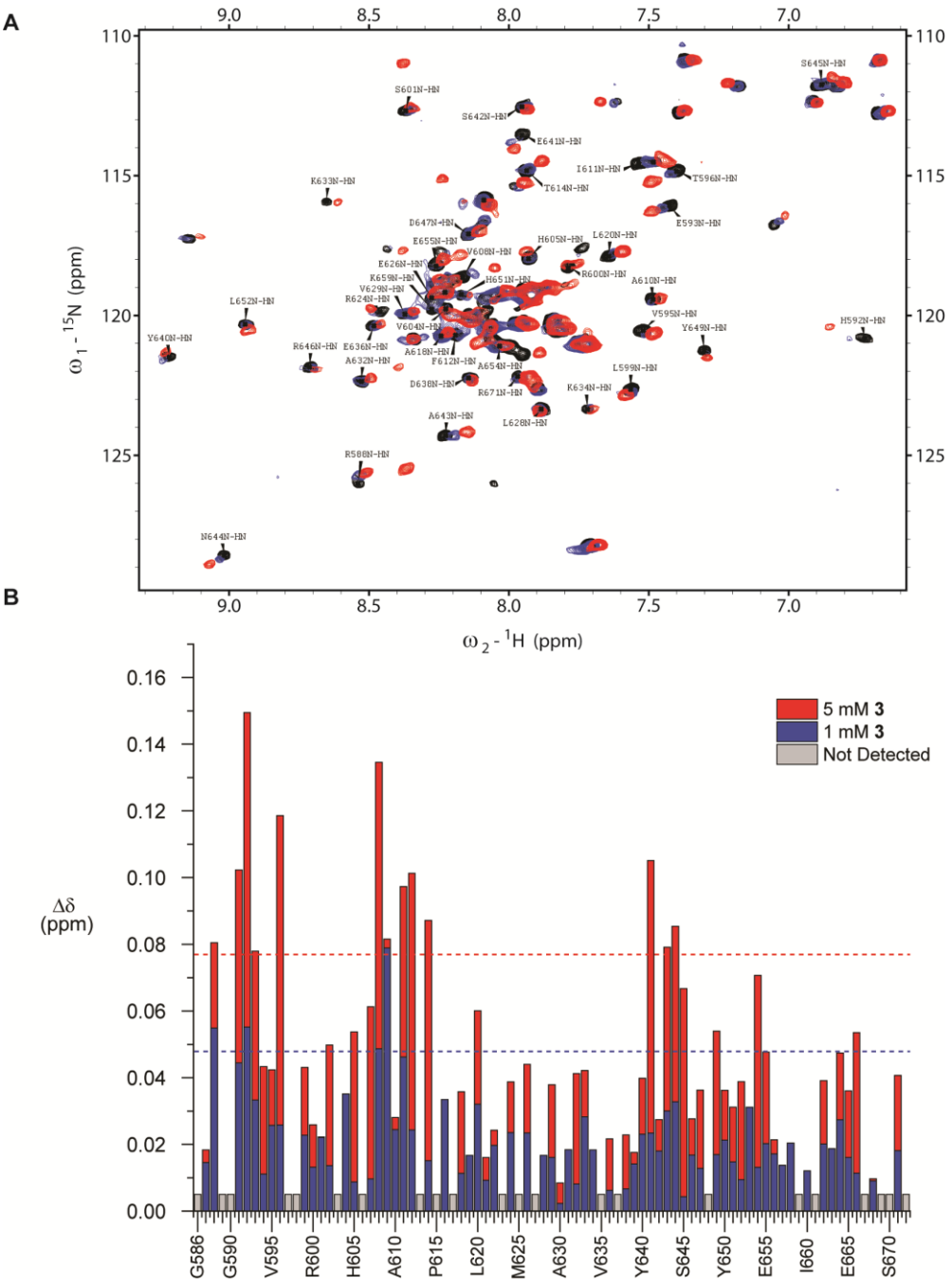


Figure 2.19. ^{1}H-^{15}N HSQC of 3 binding to KIX.
A) ^{15}N KIX alone, ^{15}N KIX + 1 mM **3**, and ^{15}N KIX + 5 mM **3** are shown in the spectrum. The peak arrows refer to KIX without ligand. B) Chemical shift perturbation mapping of KIX residues with 1 mM and 5 mM **3**. Residues that were not detected or not assigned are designated as ND. Dashed lines represent the average chemical shift perturbation + one standard deviation at the corresponding concentration. Assignments previously obtained by Majmudar et al.[42]

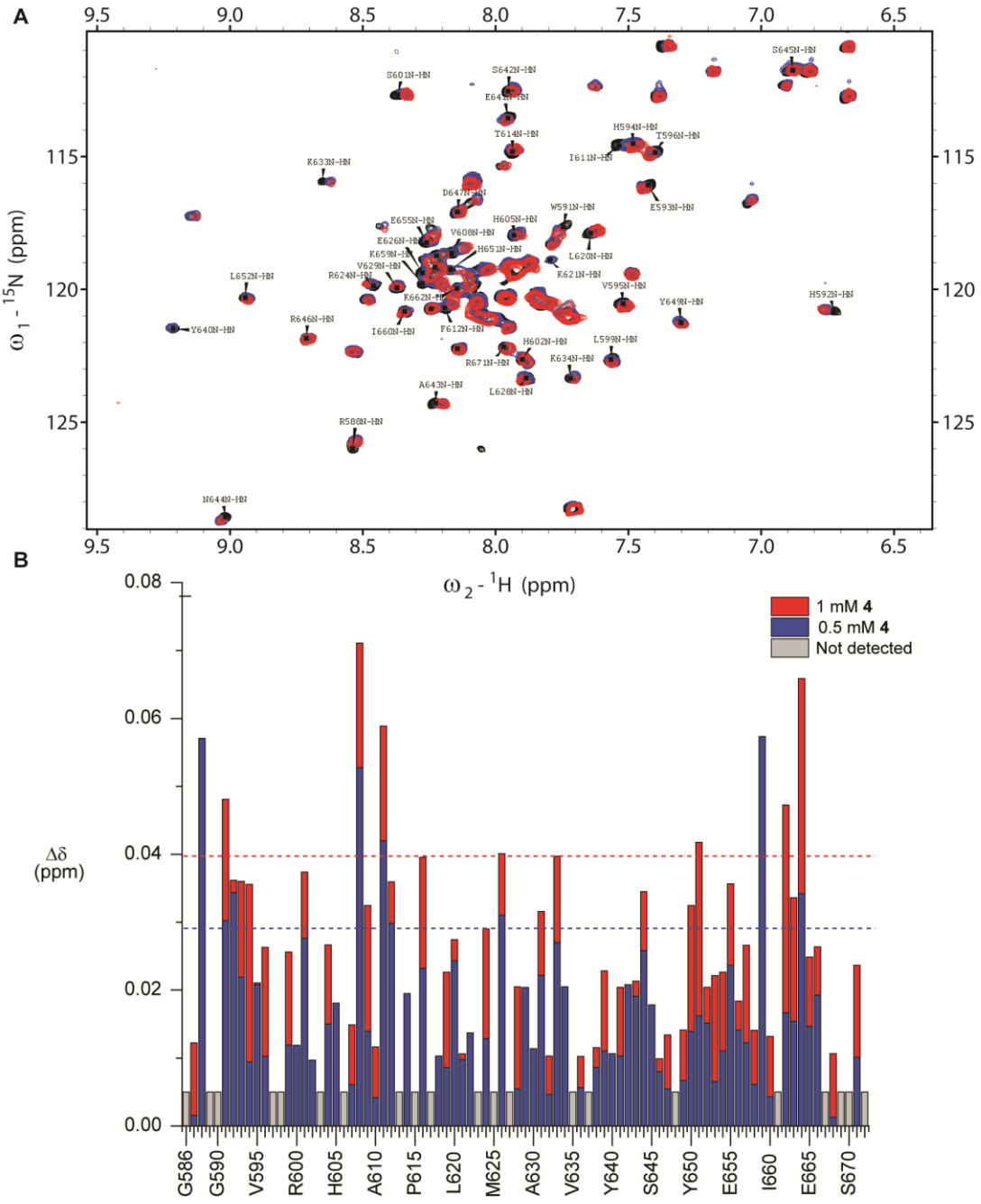


Figure 2.20. ^{1}H-^{15}N HSQC of 4 binding to KIX.
A) ^{15}N KIX alone, ^{15}N KIX + 0.5 mM **4**, and ^{15}N KIX + 1 mM **4** are shown in the spectrum. The peak arrows refer to KIX without ligand. B) Chemical shift perturbation mapping of KIX residues with 0.5 mM and 1 mM **4**. Residues that were not detected or not assigned are designated as ND. Dashed lines represent the average chemical shift perturbation + one standard deviation at the corresponding concentration. Assignments previously obtained by Majmudar et al.[42]

Consistent with PrOF NMR, perturbed residues in each case showed concentration dependent shifts and localized near the MLL binding site.[183] Additional residues were perturbed near an unstructured loop connecting helices α2 and α3 of KIX near the N-terminal portion of the protein. These perturbations may result from a conformational change induced by ligand in the MLL site rather than direct ligand binding near this loop region but will need further investigation. We conclude from this study that PrOF NMR was able to identify a related class of ligands all targetting the MLL site.

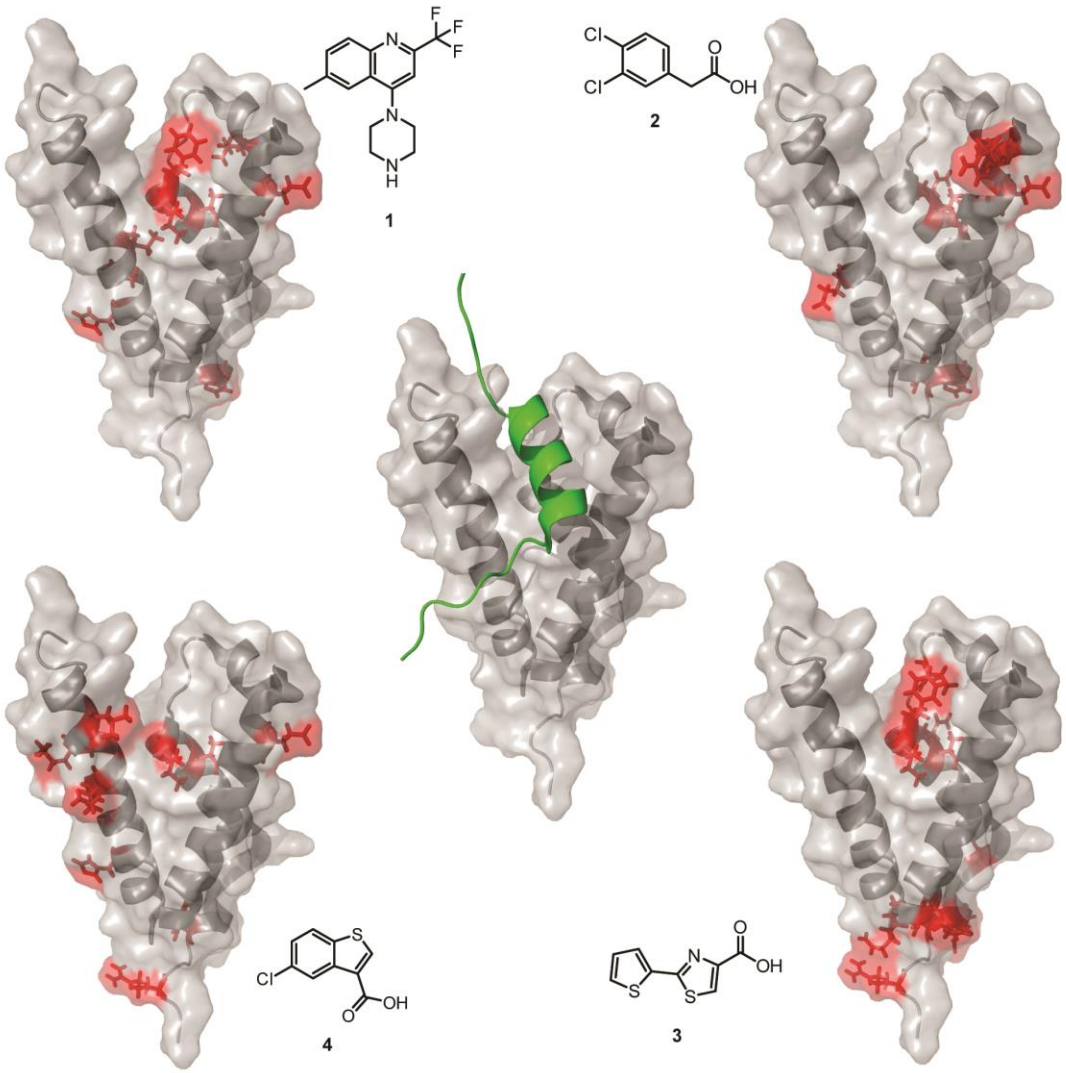
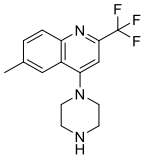
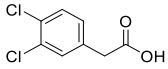
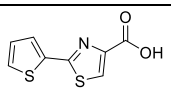
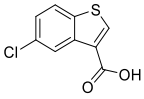
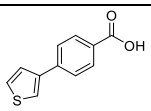
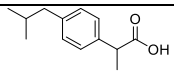
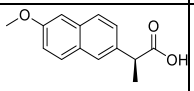


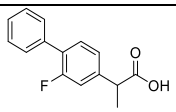
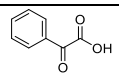
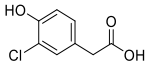
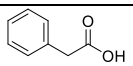
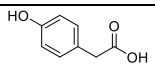
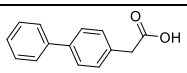
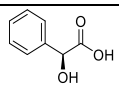
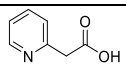
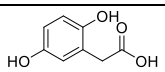
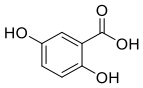
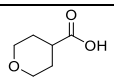
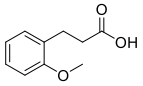
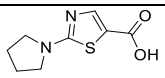
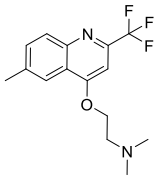
Figure 2.21. Surface map comparison of small molecule ligands from amide based HSQC data.
KIX + MLL is shown in the middle as a reference for binding site location. Each small molecule perturbs a slightly different set of residues but localize near the MLL binding site**. 4** also yields additional resonance shift affects along α3 as well near the MLL site. All four experiments also showed perturbations at residues towards the N-terminal portion of the protein and the loop between helices α2 and α3 which are thought to be the result of slight conformational changes rather than direct ligand binding. Side chains are shown for the amino acids whose amide resonance is perturbed.

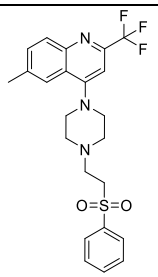
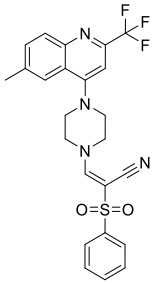
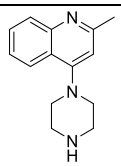
Despite few hits, the hit rates from the PrOF NMR screen could be used to assess the "druggability" of the MLL and CREB sites on KIX. Consistent with the literature, the

CREB interface is a challenging target while the MLL site is more amenable to ligand binding. For example, disulfide tethering screens reported fewer and less efficacious ligands for the CREB versus the MLL site.[186] These results support designing larger peptidomimetic inhibitors against the CREB site rather than using small molecules.[22] Both naturally occurring acidic transcription factors as well as synthetic ligands (Figure 2.1) have a higher propensity for interacting at or near the MLL region, providing a starting point for MLL-site ligand development. The KIX-interacting peptide of MLL contains the sequence DIMDFVL where both aspartic acid residues and phenylalanine are important for binding.[178] The aryl and phenylacetic acid derivatives reported here, may be minimal mimics of these motifs.

Table 2.8. Small Molecule Analysis via NMR and Computation

Compound		PrOF NMR	^{1}H-^{15}N HSQC	K_d (PrOF) (mM)	Ligand Efficiency (kcal $mol^{-1}NHA^{-1}$)	pKa	clogP
1		Yes	Yes	>10	>0.13	-	3.48
2		Yes	Yes	1.4 ± 0.3*	0.32*	4.00	2.699
3		Yes	Yes	5.3 ± 0.9	0.24	3.77	2.013
4		Yes	Yes	1.7 ± 0.6	0.29	3.67	2.720
5		Yes	-	1.6 ± 0.3	0.27	4.22	2.864
Ibuprofen		Yes	-	-	-	4.36	3.496
Naproxen		Yes	-	-	-	4.35	3.180

Flurbiprofen		Yes	-	5.9 ± 0.3	0.17	4.14	4.154
PAA1		No	-	-	-	2.20	0.865
PAA2		No	-	-	-	4.36	1.596
PAA3		No	-	-	-	4.30	2.159
PAA4		No	-	-	-	4.54	1.271
PAA5		Yes	-	6.9 ± 0.7	0.19	4.30	3.786
PAA6		No	-	-	-	3.51	0.768
PAA7		No	-	-	-	1.99	1.800
PAA8		No	-	-	-	4.54	0.426
PAA9		No	-	-	-	2.96	0.784
6		No	-	-	-	4.37	0.638
7		No	-	-	-	4.57	2.441
8		No	-	-	-	3.33	1.565
KM07508		Yes	-	-	-	-	3.589

KM09042		No	-	-	-	-	3.986
KM09000		No	-	-	-	-	3.762
9		No	-	-	-	-	2.271

2.3 **Conclusion**

PrOF NMR has several advantages as a screening method due to its sensitivity to detect changes in chemical environment including dynamic structural information of the protein, low background signal, and ease of use. Here we demonstrated the first full small molecule screen using PrOF NMR, evaluated its efficacy for druggability determination, and reported on a class of small molecules (biaryl and phenylacetic acids) as KIX ligands. Additional SAR analysis guided our investigation to discover three additional KIX-binding small molecules. The biarylacid and phenylacetic acid structural motif is corroborated by several of the reported KIX ligands (Fig. 1) providing a starting point for small molecule derivitization and development. This study also demonstrates preferential targetability for the MLL site over the CREB site. While additional optimization is required to develop the discovered ligands into potent inhibitors of KIX interactions, this study has demonstrated the applicability of PrOF NMR as a useful tool for library screening, ligand discovery, and druggability assessment. This second phase of this

project including SAR development of **2** and the installation of additional fluorine probes into KIX will be described in the next chapter.

2.4 **Methods**

Materials

Unless otherwise noted, all reagents were obtained from commercial vendors and used without further purification. For expression of proteins in *E. coli*, Lennox L. (LB) broth, unlabeled amino acids, uracil, thiamine-HCl, nicotinic acid, biotin, ampicillin, chloramphenicol, IPTG, PMSF, and Triton X-100 were purchased from RPI corp. Nucleotide bases and 3-fluorotyrosine (3FY) were purchased from Alfa Aesar. Lysozyme was purchased from Gold Biotech, HEPES was purchased from Fischer Scientific, and [^{15}N] ammonium chloride was obtained from Cambridge Isotope Laboratories. The small molecule library was acquired from Maybridge. ^{1}H, ^{13}C, and ^{19}F spectra were recorded at 500 MHz, 125 MHz, and 470 MHz respectively.

WT or 3FY-Labeled Protein Expression and Purification of His_6-KIX (586-672)

Expression

Protein expression using the same plasmid DNA for KIX was performed as previously described[41] with slight modifications. The protein expression plasmid was transformed into either the *E. coli* BL21(DE3) strain contransformed with pRARE (Novagen) (unlabeled) or DL39(DE3) cells with the plasmid pRARE (3FY-labeled) and plated on an agar plate containing ampicillin (100 mg/L) and chloramphenicol (35 mg/L). 5 mL primary cultures of DL39 or BL21 cells transformed with the KIX expression plasmid supplemented with ampicillin (100 mg/L) and chloramphenicol (34 mg/L) were grown overnight (~12-14 h, 25 °C, 250 rpm) in an incubator shaker (New Brunswick Scientific Excella E25). Primary cultures were added to fresh LB Broth (1 mL primary/50 mL media) to begin secondary culture growth. Secondary cultures were grown at (37 °C, 250 rpm) until they reached an OD_{600} of approximately 0.8 as evaluated by a Varian Cary 50 Bio UV/Vis spectrophotometer. The culture was centrifuged (8000 g, 10 min, 4 °C) to form a solid pellet of bacteria. The media was decanted, and 50-100 mL of phosphate

buffered saline (PBS) was added to resuspend the pellet. The solution was centrifuged under the same conditions and resuspended in defined media[187] containing 80-150 mg/L 3FY in the place of regular tyrosine at the original volume of LB media. A recovery time of 2-3 hours was allowed prior to the addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) to induce protein overexpression (3 h, 25 °C, 250 rpm). Cultures were spun down (6000 g, 20 min, 4 °C). Media was decanted, and the pellet was stored at -20 °C. Alternatively, overnight expression could be conducted at lower temperature (18 h, 20 °C, 250 rpm), which have resulted in the highest protein yield (70 mg/L vs 35 mg/L previously). Under these expression conditions, two other modifications to the procedure were made. The first modification involved resuspending the bacterial pellet in the defined media rather than first washing the cells with the PBS buffer. The second change was to lower the temperature in the incubator to 20 °C during the last hour of the recovery time.

Purification via Fast Protein Liquid Chromatography (FPLC)

Following cell lysis (lysis buffer, lysozyme, β-mercaptoethanol, protease inhibitor phenylmethanesulfonylfluoride (PMSF, 2 mM), and sonication), proteins were purified on a GE Äktapurifier FPLC equipped with a Superloop for large volume injections. A column containing Ni-NTA beads (Qiagen) was used for affinity purification of the His-tagged protein. An elution gradient (30-400 mM) imidazole was used to initially bind and subsequently elute the His-tagged-protein from the Ni beads. Protein was then buffer exchanged into 50 mM HEPES and 100 mM NaCl, pH 7.2 using a HiPrep 26/10 Desalting column (GE Healthcare Life Sciences) and stored at -20 °C or 4°C. During protein purification multiple aliquots of the protease inhibitor PMSF were added due to the instability of PMSF in aqueous solution.

Protein characterization

Mass spectrometry data was acquired on an Agilent MSD SL Ion Trap equipped with a liquid chromatography system (C18 column) to desalt the protein samples prior to MS analysis.

1D ^{19}F NMR.

Conditions and Parameters Used. All NMR spectra were obtained on a Bruker Avance III 500 MHz instrument. Initial measurements were obtained using a 5 mm BBFO "SmartProbe" (^{19}F S/N 550:1). The fragment screen and follow-up experiments were performed using a 5 mm Prodigy TCI Cryoprobe (^{19}F S/N 2100:1). Samples containing 30-50 µM 3FY KIX were prepared in NMR buffer as noted above with 5% D_2O and 50 µM trifluoroacetic acid (-76.5 ppm) as a reference. T_1 relaxation times were calculated using an inversion recovery experiment. Optimized parameters for PrOF NMR experiments included a 0.6 s delay time and a 0.1-0.3 s acquisition time (AQ). Spectral width and offset were 10 ppm and -136 ppm respectively. A line-broadening of 15-20 Hz was applied to all protein spectra upon processing. The screening data were acquired using 300 scans while the deconvolutions and titrations were performed with 800 scans. Data from follow-up experiments were acquired using 400-800 scans. Spectra were acquired at 294 K)

Fragment Screening. A rule of 3 (RO3) compliant library from Maybridge was obtained for fragment screening. Compounds were stored as 200 mM DMSO stock solutions. Mixtures of six compounds, consisting of one acidic, one basic, and four neutral compounds, were compiled from the library to make stock mixtures at 33.3 mM per compound. NMR samples were prepared with a total DMSO concentration of 2.5% which was shown previously to not significantly perturb the chemical shifts of the NMR spectrum. The final concentration of each fragment was 833 µM. NMR parameters are listed above.

2D ^{1}H-^{15}N HSQC NMR Experiments

Uniformly ^{15}N labeled KIX protein was expressed and purified using M9 minimal media supplemented with [^{15}N] ammonium chloride as described previously.[42] A 30-35 µM solution of ^{15}N-labeled KIX was prepared in a 9:1 H_2O:D_2O 10 mM phosphate buffer with 100 mM NaCl and 2.5% DMSO at pH 7.2. HSQC experiments were recorded at 27 °C on an Agilent/Varian 600 MHz NMR spectrometer equipped with a 5 mM Cold HCN

probe. Spectra were collected in the presence and absence of small molecule ligands discovered from the PrOF NMR fragment screen. Data was processed using NMRPipe and analyzed using Sparky (UCSF). Chemical shift changes for individual resonances were quantified using the equation $((0.2*\Delta\delta^{15}N)^2 + (\Delta\delta^{1}H)^2)^{0.5}$.

Small Molecule Extraction

Extraction of Naproxen

One CVS Naproxen Sodium liquid-filled gel capsule (220 mg (+)-1S-naproxen/capsule) was placed in water and heated to aid in dissolution. Remaining undissolved material was pelleted by centrifugation (8000 x g, 5 min, 4 °C and filtered away from the supernatant. Naproxen was precipitated out of solution using concentrated HCl. The sample was centrifuged and vacuum filtered, isolating crude naproxen. Product was recrystallized from an ethanol/dilute HCl solution and dried via vacuum filtration. Identity was confirmed by ^{1}H NMR and mass spectrometry.

^{1}H NMR (500 MHz, $(CD_3)_2CO$): δ 7.81–7.77 (m, 3 H, CH), 7.47 (dd, *J* = 8.5, 1.8 Hz, 1 H; CH), 7.29 ppm (d, ^{3}J(H,H) = 2.48 Hz, 1 H; CH), 7.16 ppm (dd, *J* = 9.0, 2.5 Hz, 1 H), 3.92 (s, 3 H), 3.90 (q, *J* = 7.1 Hz, 1 H), 1.53 (d, *J* = 7.2 Hz, 3 H): ESI-MS: calcd *m*/*z* 229.1, found 229.1 [$M^{-}H^{+}$].

Extraction of Ibuprofen

One Advil Migraine liquid-filled gel capsule (200 mg) was placed in 0.5 M NaOH and heated to aid in dissolution. Remaining undissolved material was pelleted by centrifugation (8000 x g, 5 min, 4 °C and filtered away from the supernatant. Ibuprofen was precipitated out of solution using concentrated HCl. The sample was centrifuged and vacuum filtered, isolating crude ibuprofen. Final product was recrystallized from an ethanol/dilute HCl solution. Product (racemic mixture of the R and S enantiomers) was dried via vacuum filtration and stored in a desiccator. Identity was confirmed by ^{1}H-NMR and mass spectrometry.

^{1}H NMR (500 MHz, $(CD_3)_2CO$): δ 7.27 (d, *J* = 8.3 Hz, 2 H) , 7.14 (d, *J* = 8.3 Hz, 2 H), 3.73 (q, *J* = 7.3 Hz, 1 H), 2.47 (d, *J* = 7.3 Hz, 2 H), 1.86 (sep, *J* = 6.8 Hz, 1 H), 1.44 (d, *J* = 7.1 Hz, 3 H), 0.90 (d, *J* = 6.6 Hz, 6 H): ESI-MS: calcd *m*/*z* 205.1, found 205.2 [$M^{-}H^{+}$]

Extraction of Flurbiprofen

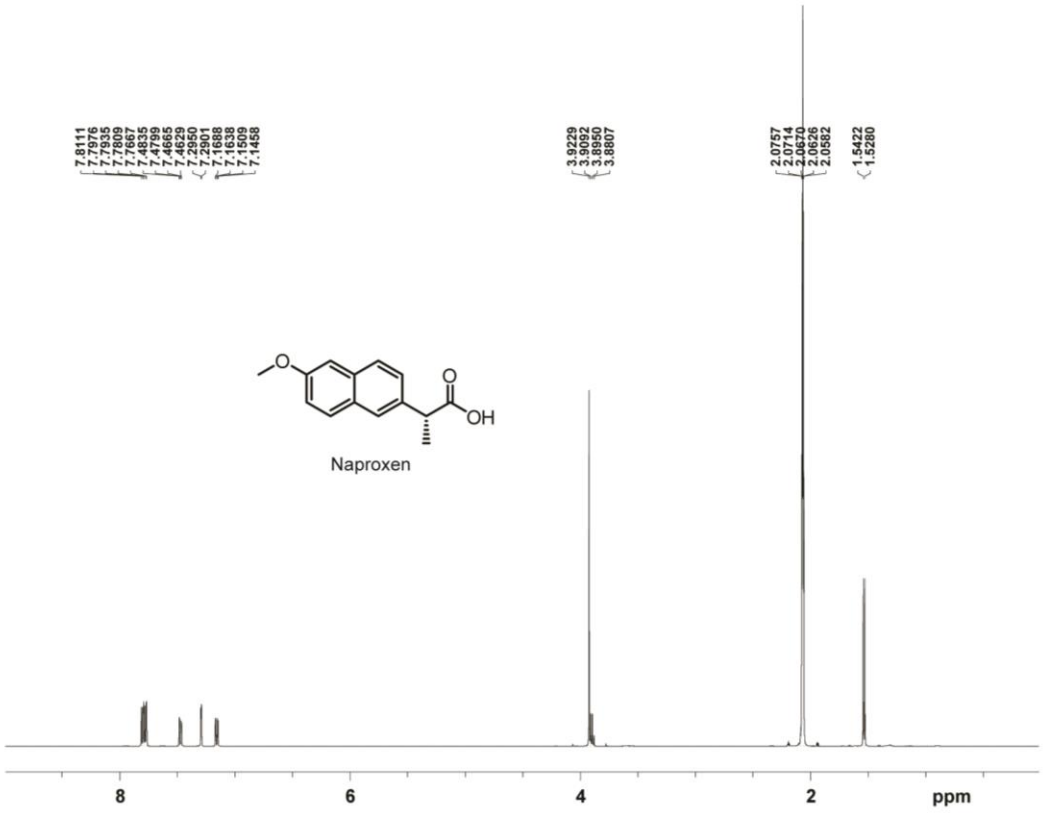
Eight Strepfen lozenges (8.75 mg flurbiprofen/lozenge) were added to 40 mL of water and heated to aid in dissolution. Flurbiprofen was precipitated using concentrated HCl. The sample was centrifuged and vacuum filtered, isolating crude flurbiprofen. The purified product was recrystallized from a methanol/water solution. The remaining aqueous solution was extracted with diethyl ether followed by an acid/base extraction to obtain a second batch of flurbiprofen. Identity was confirmed by ^{1}H NMR, ^{19}F NMR, and mass spectrometry.

^{1}H NMR (500 MHz, $(CD_3)_2CO$): δ 7.60–7.57 (m, 2 H), 7.52–7.46 (m, 3 H), 7.4 (app tt, *J* = 7.5, 1.4 Hz, 1 H), 7.26 (dd, *J* = 11.9, 1.8 Hz, 1H), 3.87 (q, *J* = 7.1 Hz, 1 H), 1.52 (d, *J* = 7.1 Hz, 3 H). ^{19}F NMR (470 MHz, $(CD_3)_2CO$) : δ -119.576 (m, 1 F), ESI-MS: calcd *m*/*z* 243.1, found 243.1 [$M^{-}H^{+}$]

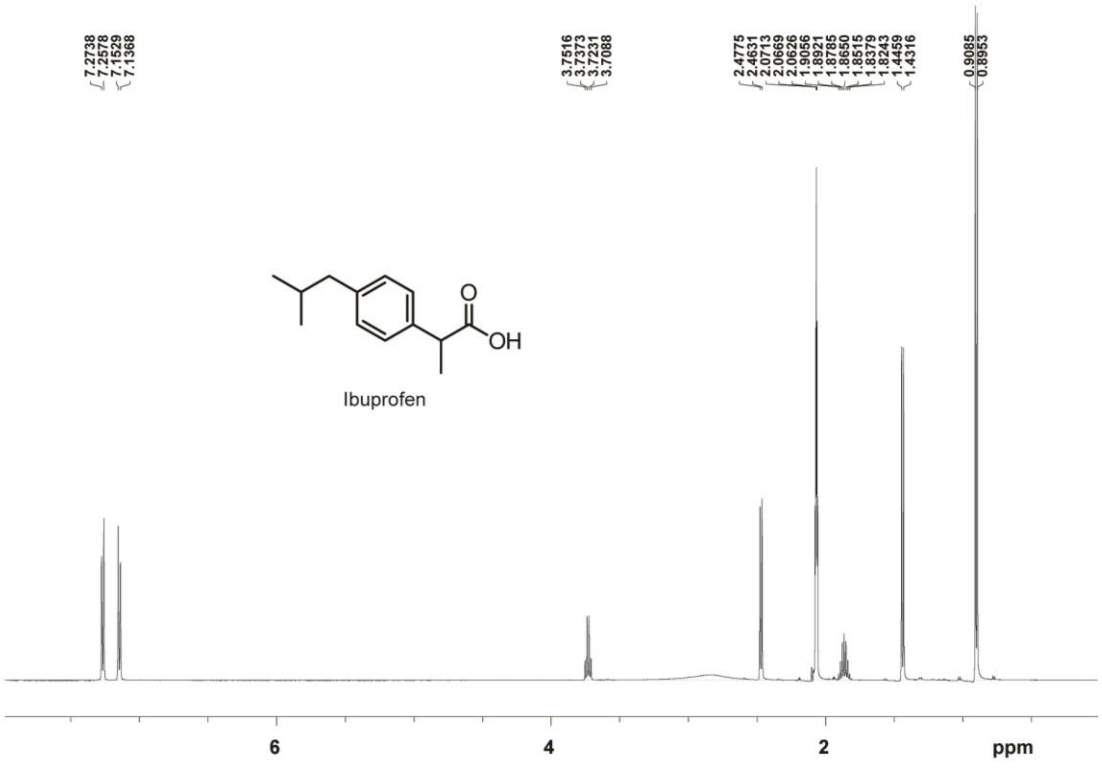
Computation

Access to the Schrodinger Software Suite was provided through the Minnesota Supercomputing Institute (MSI). SiteMap, Qikprop, and Epik modules were utilized to identify binding sites, calculate the water-partition coefficient (clogP), and pKa values respectively.

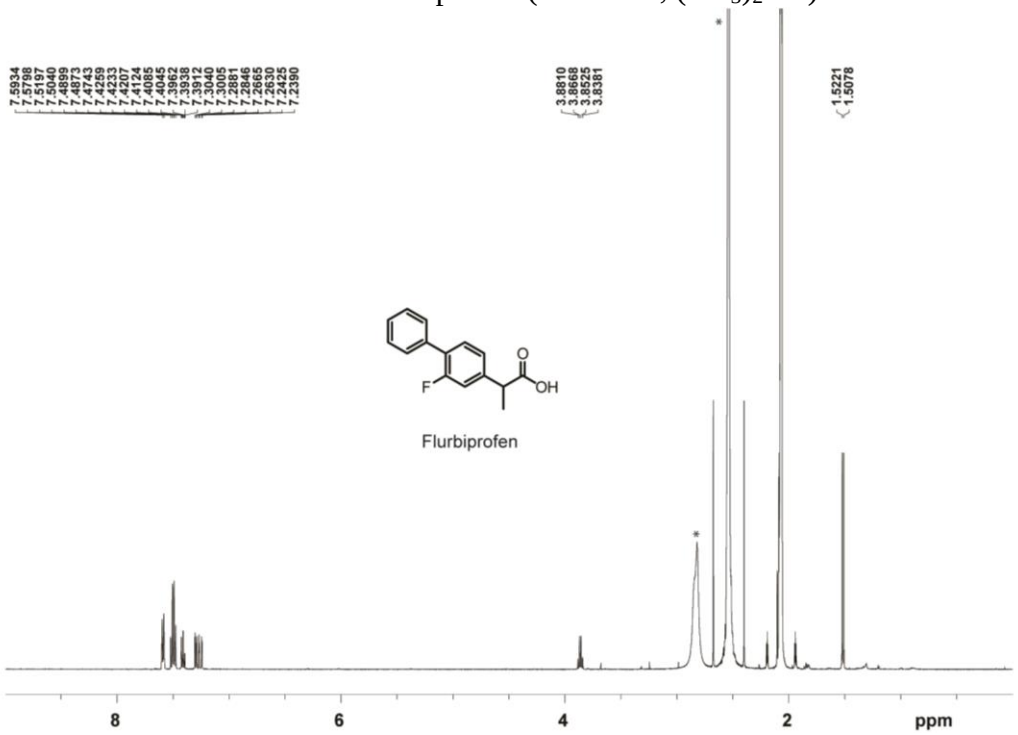
^{1}H NMR of Naproxen (500 MHz, $(CD_3)_2CO$)



^{1}H NMR of Ibuprofen (500 MHz, $(CD_3)_2CO$)

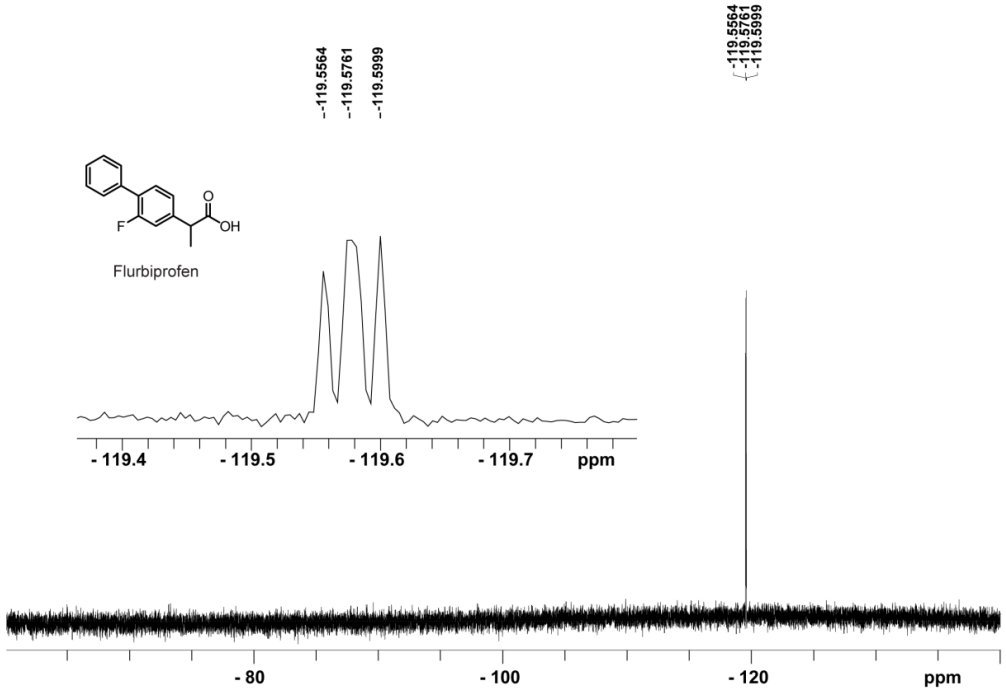


^{1}H NMR of Flurbiprofen (500 MHz, $(CD_3)_2CO$)



Stock solution of flurbiprofen (200 mM in DMSO) and analyzed in acetone-D_6 giving rise to both a water peak (2.8 ppm) and a DMSO peak (2.5 ppm) as indicated by an asterisk in addition to the acetone solvent peak.

^{19}F NMR of Flurbiprofen (470 MHz, $(CD_3)_2CO$)



Chapter 3. Dual Labeling of the CBP/p300 KIX domain for 19F NMR leads to identification of a new small molecule binding site

This paper is manuscript was resubmitted August 7th, 2017 as follows:

"Dual labeling of the CBP/p300 KIX domain for 19F NMR Leads to identification of a new small molecule binding site," C. T. Gee*, K. E. Arntson*, E. J. Koleski, W. C. K. Pomerantz,

Motivation: This manuscript was the follow-up to the previous chapter and the combination of two different research directions. The first goal was to take one of the hit compounds from the fragment screen (as discussed in Chapter 2) and develop it into a more potent small molecule inhibitor. The second goal was to push the limits of fluorine labeling in a protein. We already had established the process for labeling a protein with a single type of fluorinated amino acid, so we wanted to explore the feasibility of incorporating more than one type of amino acid in the same protein. These two projects came together when one of the developed ligands demonstrated some unexpected behavior by not perturbing the newly incorporated 4-fluoro-phenylalanine, leading to the discovery of the new binding site.

3.1 **Introduction**

Small-molecule inhibition of protein-protein interactions (PPIs), while successful in several cases,[7-8, 13] is still a challenging area of research.[169] Due to their large surface areas, multiple low-affinity binding partners in the case of transcription factors, and allosteric networks, PPIs are notoriously difficult to target with small molecules.[14] In contrast with enzymes (e.g., kinases)[188] or epigenetic "reader" proteins (e.g., bromodomains)[189] that typically have one primary binding site, transcription factors involved in PPIs often have multiple native binding partners which in some cases, bind to different sites on a protein surface.[15] These sites may be allosterically coupled. Multiple

binding sites can make it difficult to identify exactly where a small molecule is interacting with a protein. For example, the KIX domain of the CREB Binding Protein (CBP/p300) interacts with over a dozen different transcription factors through two allosterically coupled sites.[190] Here we report on a new small-molecule binding site distinct from those used by native transcription factors. (Figure 3.1).

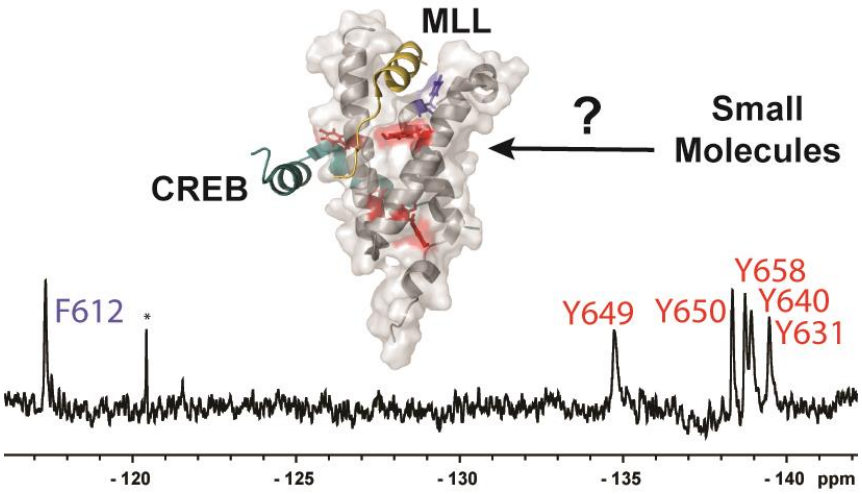


Figure 3.1 New small molecule binding site on KIX discovered through the use of a dual-fluorine-labeled protein.

In a recent characterization of KIX, Pomerantz et al. found that 3-fluorotyrosine incorporation at all five tyrosine sites for Protein-Observed Fluorine NMR spectroscopy (PrOF NMR) was sufficient for distinguishing the binding footprints of three native ligands from the transcriptional activation domains of c-Myb, MLL, and CREB.[41] These proteins interact with the KIX domain through two distinct binding sites. Allosteric effects on the protein were further shown to be different between c-Myb and CREB despite their similar binding surfaces on KIX.[191-192] Screening of 508 small molecules against this fluorinated protein by PrOF NMR revealed several molecules which perturbed the resonance for Y631 consistent with binding near the second transcription factor site engaged by MLL.[193]

We and others have shown that fluorine-labeling of aromatic amino acids for PrOF NMR spectroscopy provides structural information on the ligand-binding site and can quantify interaction affinities.[74, 194-195] Due to the speed of this method, which can be run

in five minutes or fewer at low- to mid- micromolar concentrations, it has been used for screening libraries of low molecular weight compounds called fragments in both primary and follow-up screens on small- to medium-sized proteins.[75, 99, 193, 196] Paramagnetic additives have further increased the throughput of the method, reducing experiment times by 60%.[107]

Although enriched at protein-protein interfaces, aromatic amino acids occur at low frequency in proteins.[170] Thus this sparse labeling strategy may not be sufficient to elucidate the protein binding site of small molecule ligands in all cases. For example, in the MLL binding site, the presence of a single tyrosine residue, Y631, limits the amount of structural information that can be inferred and increases the likelihood of false negatives in screening. However, there is a phenylalanine present in the MLL binding site near Y631, which could also serve as an additional probe for PrOF NMR if simultaneously fluorine labeled (Figure 3.2).

Budisa and co-workers were the first to show the ability to incorporate multiple non-natural amino acids into proteins via sequence selective methods using auxotrophic bacteria for 4-fluoroproline, 6-fluorotryptophan, and 4-fluorophenylalanine.[197] Fluorinated aromatic amino acids are particularly challenging due to their toxicity towards bacteria during recombinant protein expression conditions. In these prior studies, labeling was high but still resulted in a heterogeneously labeled sample, which may have a significant effect on the resulting NMR spectrum.

Due to the challenges presented by the bacteriostatic effects of fluorinated amino acids on bacterial cell growth,[198] a fluorine NMR study on a multi-labeled protein has yet to be reported. To address this need, we explore a dual-labeling strategy using both 3-fluorotyrosine (3FY) and 4-fluorophenylalanine (4FF) as a test case for characterizing native peptides sequences and newly discovered small molecules for the KIX domain for future chemical probe development. We report on a new small molecule and our data supports a new binding site, distinct from native transcription factor binding sites for regulating KIX PPIs. Results from this study provide insights into new approaches for using fluorine-labeled proteins to characterize protein-ligand interactions by PrOF NMR.

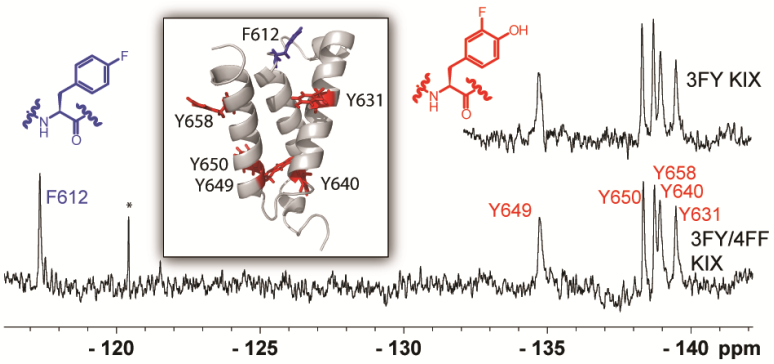


Figure 3.2 PrOF NMR comparison of 3FY KIX and 3FY/4FF KIX.
Top Right: PrOF NMR spectrum of 3FY KIX. Bottom: PrOF NMR spectrum of 3FY/4FF KIX. Inset: Solution structure of KIX with tyrosines and phenylalanine shown as sticks. (PDB ID: 1KDX) *This resonance is an artifact that does not originate in the sample.

3.2 **Results and Discussion**

To obtain a protein labeled with two different fluorinated amino acids, 3FY and 4FF, an auxotrophic cell needed to be selected which suppressed natural production of both tyrosine and phenylalanine. While the herbicide glyphosate inhibits the shikimate pathway in bacteria and thus suppresses tyrosine and phenylalanine biosynthesis to induce auxotrophy, this method results in low protein expression yields and high costs due to the amount required in the expression (1 g/L expression).[199-200] Alternatively, the DL39(DE3) cell line is auxotrophic for both tyrosine and phenylalanine.[200] We have obtained yields as high as 70 mg/L of singly-labeled KIX with 3FY with this cell line, and thus chose it for our study.[41, 193] Previous expressions for 4FF-labeled KIX resulted in highly variable yields, possibly due to varying conditions, including expression and recovery time (the amount of time after the media change for bacteria to recover). Of note, we observed significant bacterial growth inhibition in the case of incorporating 4FF where yields were lower than an independent expression of KIX with 3FY. It has previously been demonstrated by Bann et al. that DL39(DE3) cells give higher 4FF-protein yields when supplied with 0.1 mM phenylalanine.[200] We thus used this approach in our first set of optimization conditions.

The optimization of 4FF KIX and by extension 3FY/4FF KIX expression, is therefore contingent on two conflicting variables. Higher unlabeled phenylalanine concentration added to expression media, increases yield but hinders fluorine-labeling, while lower concentration of phenylalanine in the media decreases yield and increases fluorine-labeling. Optimization experiments identified ideal conditions for expression: media containing 10 μM phenylalanine, 160 μM 4FF, and a 30 min recovery time after the defined media change, resulting in 80% fluorine labeling (Table 3.1). Following the optimization process, expression yields of 4FF KIX ranged from 15-22 mg/L with 80-88% fluorine incorporation.

Table 3.1 Optimization of 4FF KIX Conditions

4FF KIX			
4FF (μM)	F (μM)	Recovery Time (min)	% Fluorine Labeling
160	160 80 40 0	30 min	32% 27% 28% 84%
160	40 20 0	30 min	33% 66% 75%
640 320	40	30 min	36% 27%
160	20 10 0	0 min	21% 32% 66%
160	20 10 0	30 min	73% 80% 75%
160	10 7.5 5	40 min	73% 84% 79%
160	10	60 min 120 min	68% 77%
3FY/4FF KIX			
160	0	40 min	95.5%

The first successful expression of a dual-labeled 3FY/4FF KIX was achieved using the conditions established for 4FF KIX above, with the addition of 400 μM 3FY and removal of tyrosine from the defined media. High fluorine incorporation (85%) was achieved, and a ^{19}F NMR spectrum of the dual-labeled sample was obtained. Under these conditions, yields of 3FY/4FF KIX ranged from 14-22 mg/L, with fluorine incorporation ranging from 70-85%. Although this degree of incorporation may be sufficient in some cases, we ultimately wanted to obtain higher incorporation levels. By leaving phenylalanine out of expressions, fluorine incorporation of 95% can reliably be obtained, reducing protein yield to 3-7 mg/L. All subseqent studies have used this highly-labeled protein (Figure 3.2, Figure **3.3**).

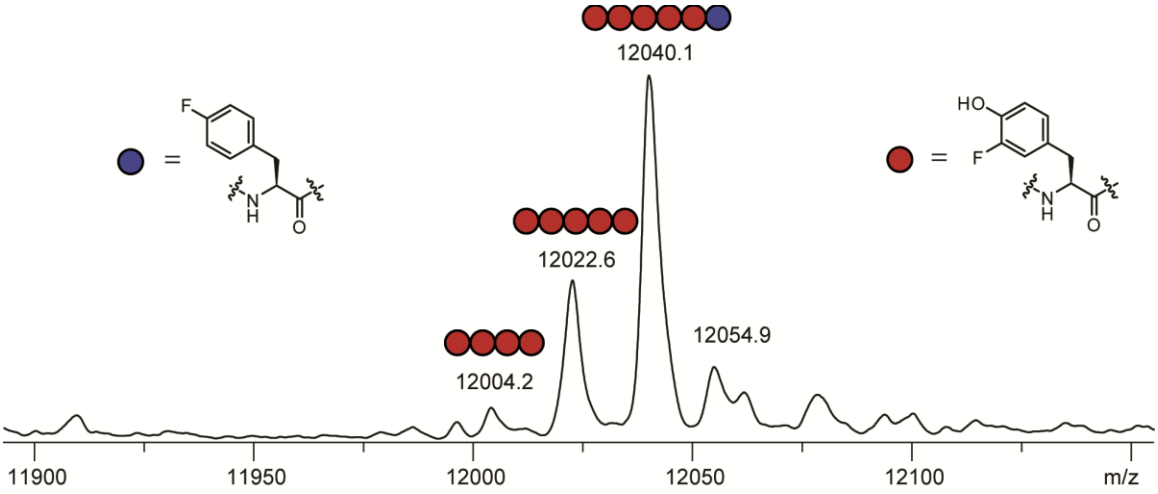
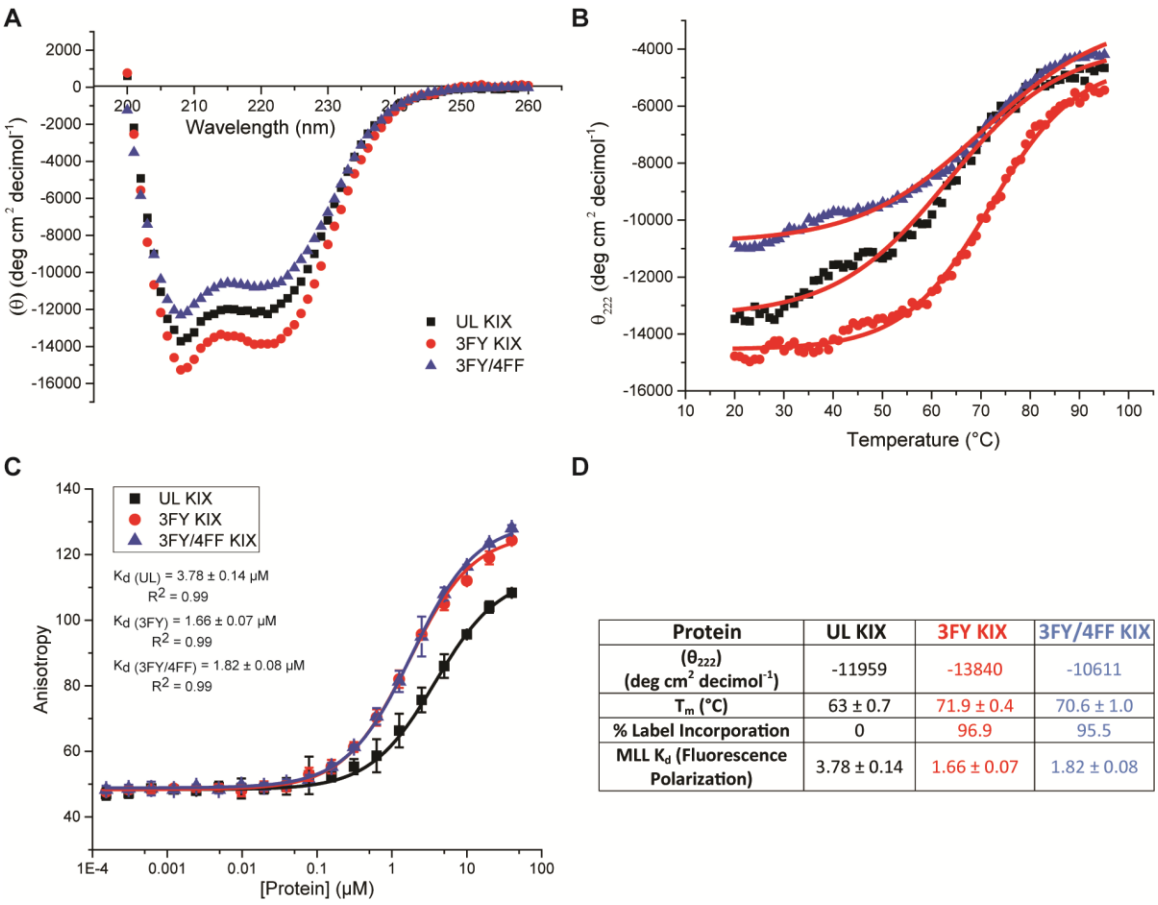


Figure 3.3 The deconvoluted ESI-mass spectrogram of a 3FY/4FF KIX sample demonstrating high fluorine incorporation.
The peak at 12004.2 is a 67% labeled protein, the peak at 12022.6.6 is an 83% labeled protein, the peak at 12040.1 is the 100% labeled protein containing all six fluorinated residues, and the peak at 12054.9 is consistent with a small percentage of an oxidized product that was presumed to occur from methionine oxidation. Percent labeling is calculated with the equation below, using the integration values of the various deconvoluted mass peaks as the values for F KIX.

$$\%\ Incorporation = \frac{(0F\ KIX) * 0 + (1F\ KIX) * 1 + (2F\ KIX) * 2 \ldots (6F\ KIX) * 6}{(0F\ KIX) * 6 + (1F\ KIX) * 1 + (2F\ KIX) * 6 \ldots (6F\ KIX) * 6}$$



Protein	UL KIX	3FY KIX	3FY/4FF KIX
(θ_{222}) (deg cm^2 decimol^{-1})	-11959	-13840	-10611
T_m (°C)	63 ± 0.7	71.9 ± 0.4	70.6 ± 1.0
% Label Incorporation	0	96.9	95.5
MLL K_d (Fluorescence Polarization)	3.78 ± 0.14	1.66 ± 0.07	1.82 ± 0.08

Figure 3.4 Structure and function validation of dual labeled protein.
A) CD Spectra, B) thermal melt, and C) MLL affinity comparisons of unlabeled (UL), 3FY-labeled, and 3FY/4FF-labeled KIX. Unlabeled protein is plotted with squares. 3FY protein is plotted with circles. 3FY/4FF protein is plotted with triangles. D) Summary table of structural and functional data.

Once protein expression conditions were established, we first sought to test the possibility of any significant structural or functional perturbation from fluorine incorporation. In prior reports, fluorescence polarization (FP) binding studies and circular dichroism (CD) investigations of thermal stability and secondary structure revealed minor perturbation of structure and binding from 3FY incorporation. Far-UV CD spectra showed all three protein constructs to be highly α-helical based on molar ellipticity at 222 nm, with the dual-labeled protein with an 11% lower molar ellipticity (Figure 3.4A, D). Similar to 3FY-labeling, the dual-labeled protein did show a remarkable thermal stabilization of 7 degrees relative to the unlabeled protein. The origin of this effect is

unclear. Wang et al. demonstrated that a covalent ligand was able to stabilize KIX by 15-18 degrees leading to the first and only x-ray structure of this protein.[47] We speculate that fluorine labeling may serve the same purpose. Because we labeled F612 which is in the MLL binding site, we also compared binding of the unlabeled, 3FY-labeled, and 3FY/4FF dual-labeled-KIX protein to MLL by FP. We saw approximately a 2-fold difference in binding affinity across samples (Figure 3.4C, D, K_d = 3.78 μM, 1.66 μM, and 1.82 μM, respectively). Together these results support only a modest perturbation from fluorine labeling to global secondary structure and ligand binding.

The fluorine nucleus is highly sensitive to small changes in chemical environment and thus is also a sensitive reporter of subtle conformational effects.[74] In our first PrOF NMR experiment on this new construct, we compared the chemical shift of the resonances of our 3FY-labeled KIX protein to the 95% 3FY/4FF- labeled protein (Figure 3.2). In both cases, the 3FY resonances were very similar (0.04 ppm average difference) between the two spectra indicating a minimal change in chemical environment despite the slightly lower molar ellipticity from CD. To date, this is the first example of dual-labeled protein by ^{19}F NMR. The single 4FF resonance found at -117.3 ppm is also in a similar region as our previously reported singly-labeled protein.[41] The large chemical shift dispersion between the two sets of resonances (~20 ppm between 3FY and 4FF) indicates distinct windows for fluorinated aromatic amino acid resonances. Furthermore, 5-fluorotryptophan resonances typically originate near -125 ppm, allowing sufficient chemical shift space for future triple-labeling experiments.[195]

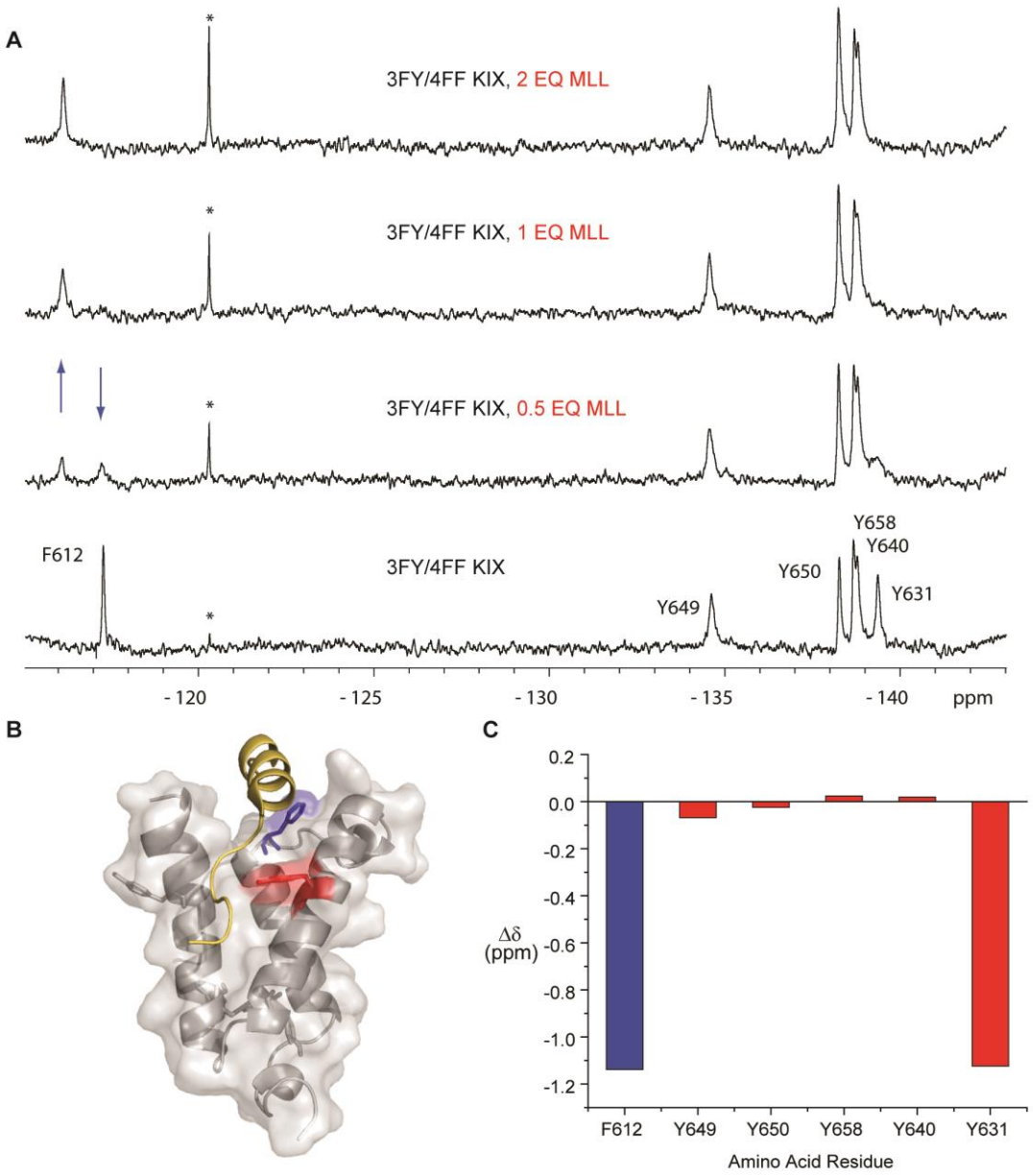


Figure 3.5 PrOF NMR results with 3FY/4FF KIX and MLL.
A) Stacked PrOF NMR spectra of 3FY/4FF KIX and increasing concentrations of MLL. Y631 and F612 are the only two resonance that are perturbed. MLL also demonstrates slow exchange kinetics, as noted by the unbound and bound states being resolved at 0.5 EQ MLL. * This resonance is an artifact that does not originate in the sample. B) Structure of KIX:MLL complex (2AGH) with Y631 and F612 highlighted on the surface. Both are in close proximity to the native MLL peptide. C) Chemical shift perturbations of 3FY/4FF KIX with 2eq MLL.

Table 3.2 3FY/4FF KIX ^{19}F NMR chemical shift perturbations at varying MLL concentrations

[MLL]	Chemical shift (ppm)	**F612**	**Y649**	**Y650**	**Y658**	**Y640**	**Y631**
0	δ (ppm)	-117.2887	-134.6191	-138.2642	-138.6706	-138.7750	-139.3630
2 EQ	Δδ (ppm)	-1.1385	-0.0680	-0.0250	0.0236	0.0189	-1.1238

With both sets of fluorinated resonances clearly resolved, we sought to characterize the binding interactions of native and synthetic ligands. The solution NMR structure of MLL binding to KIX (PDB ID: 2AGH) indicates close interactions with the MLL transcriptional activation domain and KIX at both F612 and Y631. Our PrOF NMR experiment is able to capture these interactions, where both resonances yield large perturbations (Δδ = 1.1-1.2 ppm) upon increasing ligand concentration in the slow exchange regime (Figure 3.5). A second transcription factor, CREB, binds a second site on the KIX surface but is known to enhance MLL binding through an allosteric network.[192] We previously showed that the 3FY resonances (Y658, Y650, and Y649) which mapped to the binding site were affected as well as Y631 in the MLL binding site. Independent experiments with 4FF-labeled KIX showed no effects on the F612 resonance.[41] Because the singly-labeled KIX and dual-labeled KIX are different and could be presenting slightly different binding surfaces, we characterized this binding-induced allosteric effect again with dual-labeled KIX. (Figure 3.6) In this experiment, Y631 was similarly perturbed and F612 was unaffected, supporting an allosteric network that affects Y631, but does not involve F612 within the binding site.

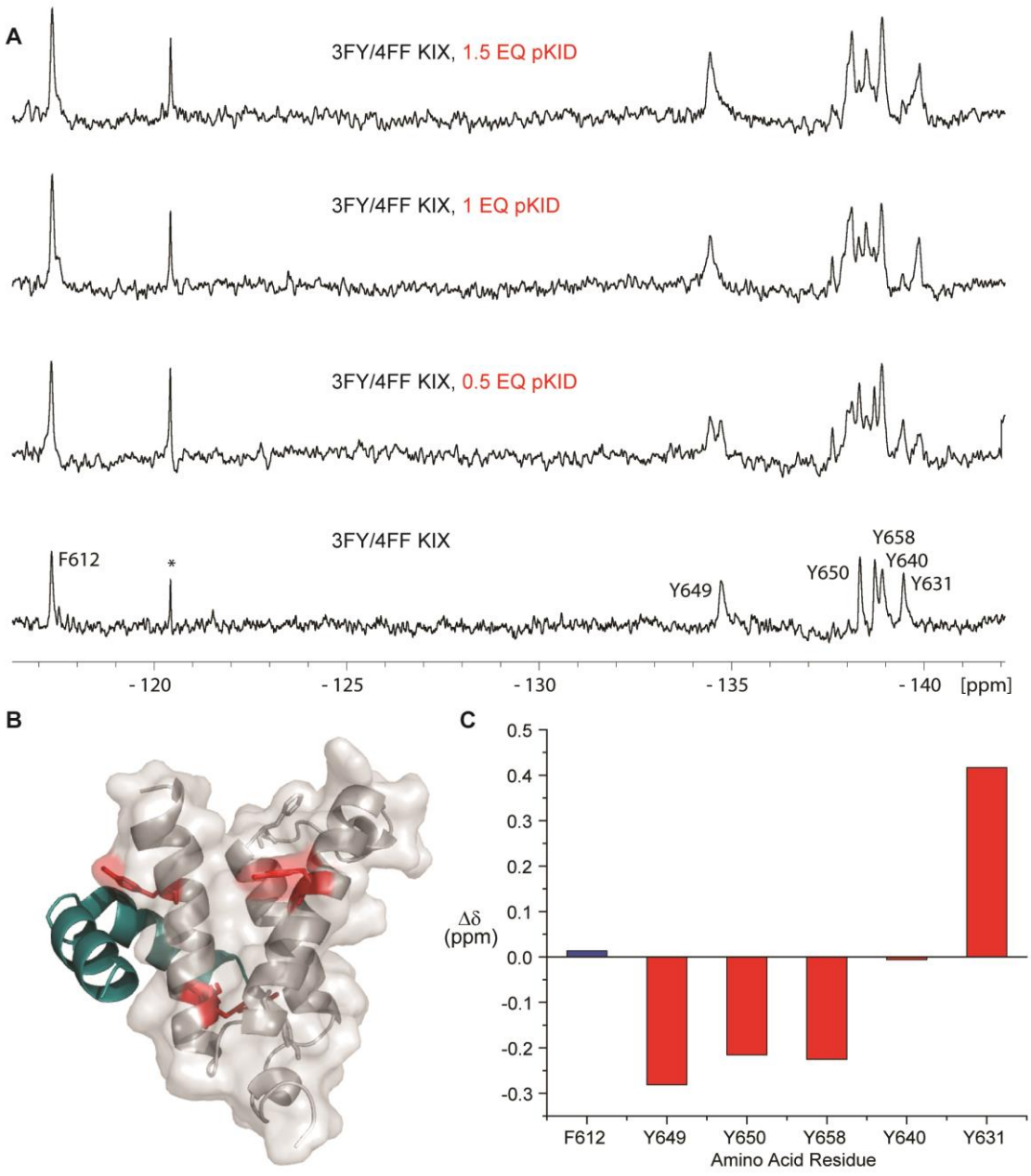


Figure 3.6 PrOF NMR results with 3FY/4FF KIX and pKID.
A) Stacked PrOF NMR spectra of 3FY/4FF KIX and increasing concentrations of pKID. Y649, Y650, and Y658 are perturbed by direct contact while Y631 is allosterically perturbed. * This resonance is an artifact that does not originate in the sample. B) Structure of KIX:MLL complex (1KDX) with Y649, Y650, Y658, and Y631 highlighted on the surface. C) Chemical shift perturbations of 3FY/4FF KIX with 1.5 eq pKID.

Table 3.3 3FY/4FF-KIX ^{19}F NMR chemical shift perturbations at varying pKID concentrations.

[pKID]	Chemical shift (ppm)	**F612**	**Y649**	**Y650**	**Y658**	**Y640**	**Y631**
0	δ (ppm)	-117.3396	-134.7238	-138.3410	-138.7253	-138.9210	-139.4710
1.5 EQ	Δδ (ppm)	0.0136	-0.2809	-0.2155	-0.2253	-0.0061	0.4167

Few small molecules have been reported to bind to the KIX domain. Naphthol AS-E phosphate is one of the most well characterized PPI inhibitors of KIX-transcription factor interactions between both c-Myb and CREB.[36, 201] Our studies and those by Best et al. using ^{1}H-^{15}N HSQC, argue against binding at the CREB or c-MYB site, despite the similarity of the phosphate group in the inhibitor and phosphoserine 133 of CREB which binds to KIX, in contrast to the binding-site hypothesis proposed by Uttarkar et al.[201] Similar to our experiments with CREB, we were concerned that KIX labeled with 4FF might be altering the protein structure and masking interactions within the MLL binding site. Titrating naphthol-AS-E phosphate against the dual-labeled protein however, confirmed that Y631 is the sole residue that exhibits a large dose dependent effect and supports the existence of an alternate small molecule binding site distinct from those used by the native transcription factors (Figure 3.7).

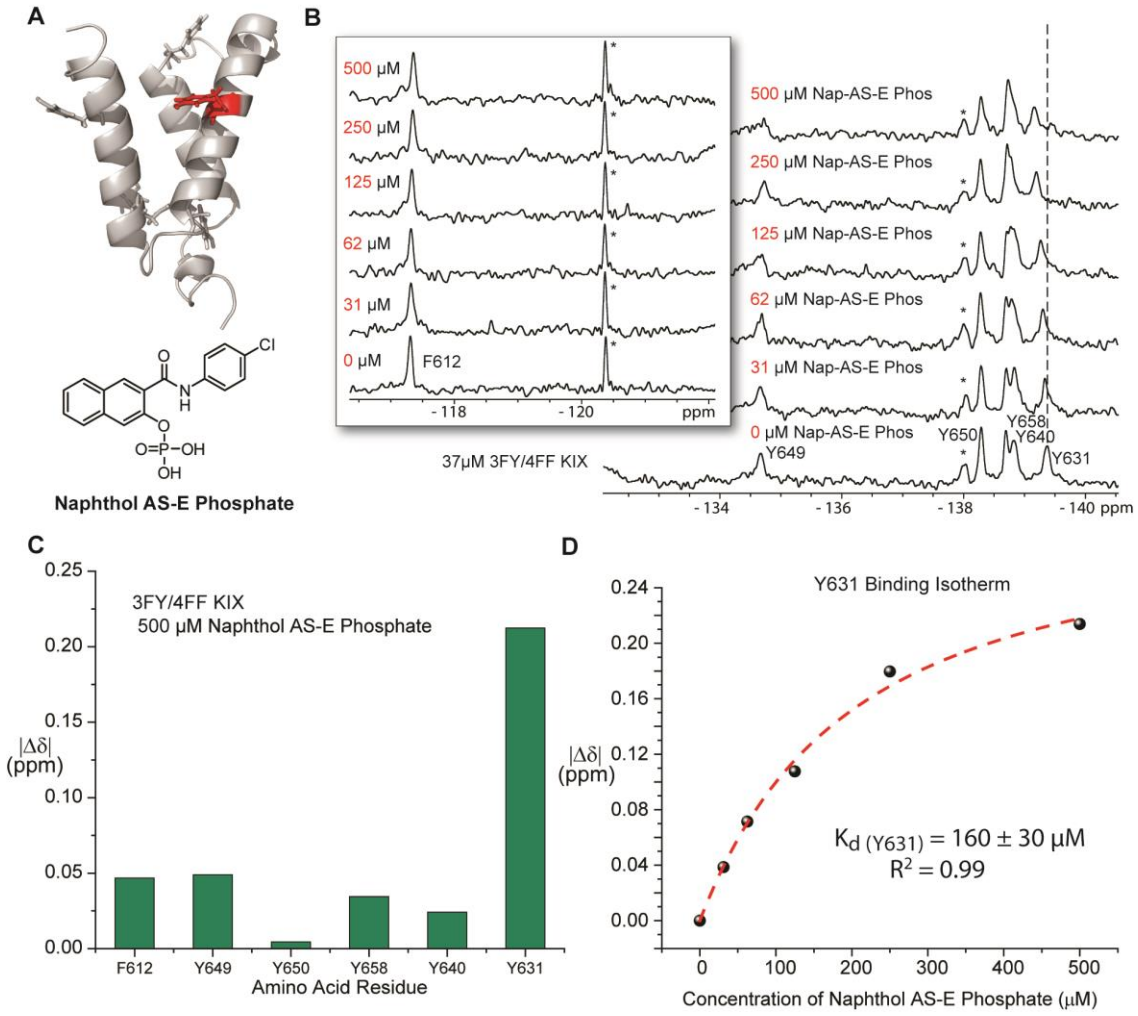


Figure 3.7 ^{19}F NMR spectral analysis of 3FY/4FF KIX in the presence of increasing concentrations of naphthol AS-E phosphate.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. * This resonance is an artifact that does not originate in the sample. B) Stacked spectra of 3FY/4FF KIX with increasing concentrations of naphthol AS-E phosphate. Inset) Resonances from F612. C) Absolute value of chemical shift perturbations for all 3FY/4FF KIX resonances at 500 µM naphthol AS-E phosphate. D) Binding isotherm of Y631 perturbation for the titration of naphthol AS-E phosphate.

Table 3.4 3FY/4FF-KIX ^{19}F NMR chemical shift perturbations at varying naphthol AS-E phosphate concentrations

[Nap]	Chemical shift (ppm)	F612	Y649	Y650	Y658	Y640	Y631
0	δ (ppm)	-117.311	-134.6720	-138.2915	-138.6966	-138.8242	-139.3750
31.25	|Δδ| (ppm)	0.0084	0.0065	0.0022	0.0060	0.0106	0.0386
62	|Δδ| (ppm)	0.0090	0.0189	0.0114	0.0043	0.0351	0.0714
125	|Δδ| (ppm)	0.0225	0.0220	0.0101	0.0794	0.0482	0.1076
250	|Δδ| (ppm)	0.1114	0.0593	0.0119	0.0202	0.0552	0.1796
500	|Δδ| (ppm)	0.0467	0.0490	0.0045	0.0344	0.0242	0.2138

The above results led us to test small molecule **1**, discovered in our fragment screen, which affected Y631 and at higher concentrations exhibited a non 1:1 interaction.[193] We speculated that this interaction was caused by two molecules fitting within the MLL site. However, addition of **1** at increasing concentration did not affect F612. Either **1**, due to its small size, was too far away from F612, or it bound near but outside the MLL binding site similar to naphthol AS-E-phosphate. To this end, we designed several more extended molecules through acylation with amino acids tryptophan, phenylalanine, and tyrosine (molecules **2, S3, S4**). Structure-activity relationships studies showed the carboxylic acid of **1** to be important for binding, as noted by the methyl ester of **1** being inactive (Figure 3.8).

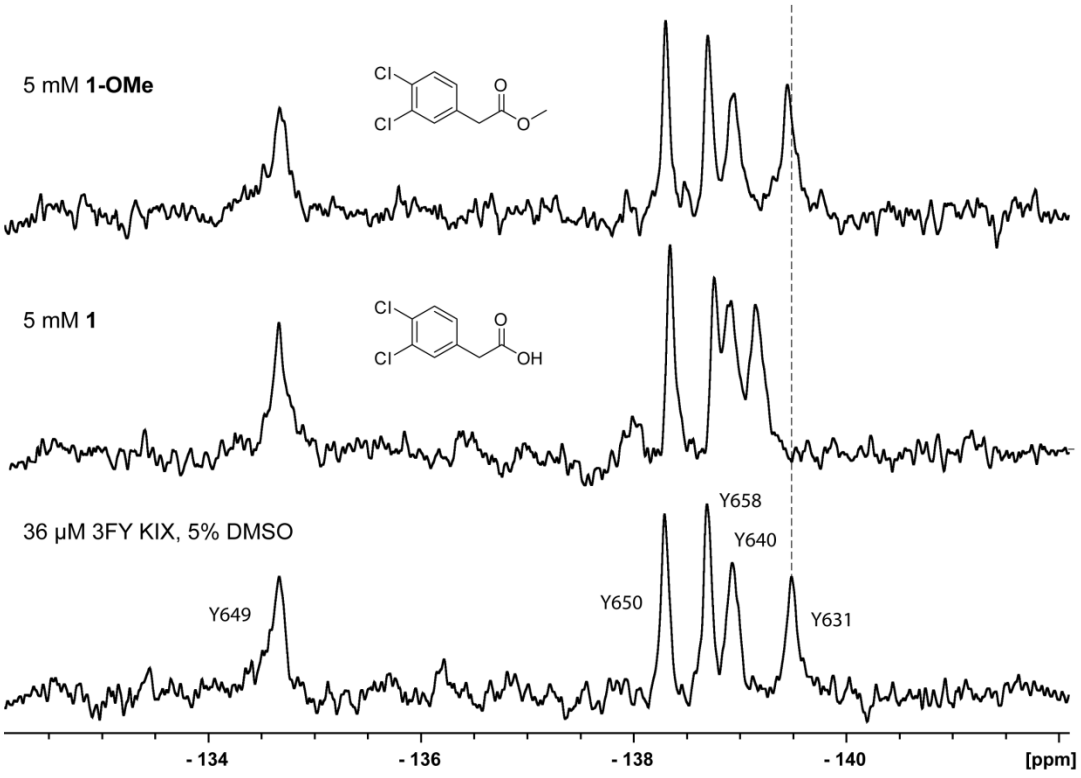


Figure 3.8 PrOF NMR data comparing 1 and 1-OMe.
Tthe methyl ester eliminates the activity of the parent compound based on the observation that Y631 is no longer perturbed.

Therefore, **2**, **S3**, and **S4** were synthesized to maintain the carboxylic acid. Titration of **2**, against 3FY KIX indicated binding near the MLL site, and a saturating binding isotherm, consistent with 1:1 binding was observed, yielding a K_d of 740 µM (Figure 3.9). **S3** and **S4** were also titrated against 3FY KIX, but there was little to no improvement in affinity relative to the parent compound, **1** (Figure 3.10, Figure 3.11). Testing **2** with the 3FY/4FF labeled protein**,** again perturbed Y631 near the MLL binding site but did not affect F612 (Figure 3.14, Figure 3.12). The profile of **2** binding to both 3FY KIX and 3FY/4FF KIX indicated similar binding modes though a slightly weaker affinity (Figure 3.12, Figure 3.13 A,B). Together these data support **1, 2,** and naphthol AS-E phosphate interacting at a similar binding site, distinct from native transcription factors.

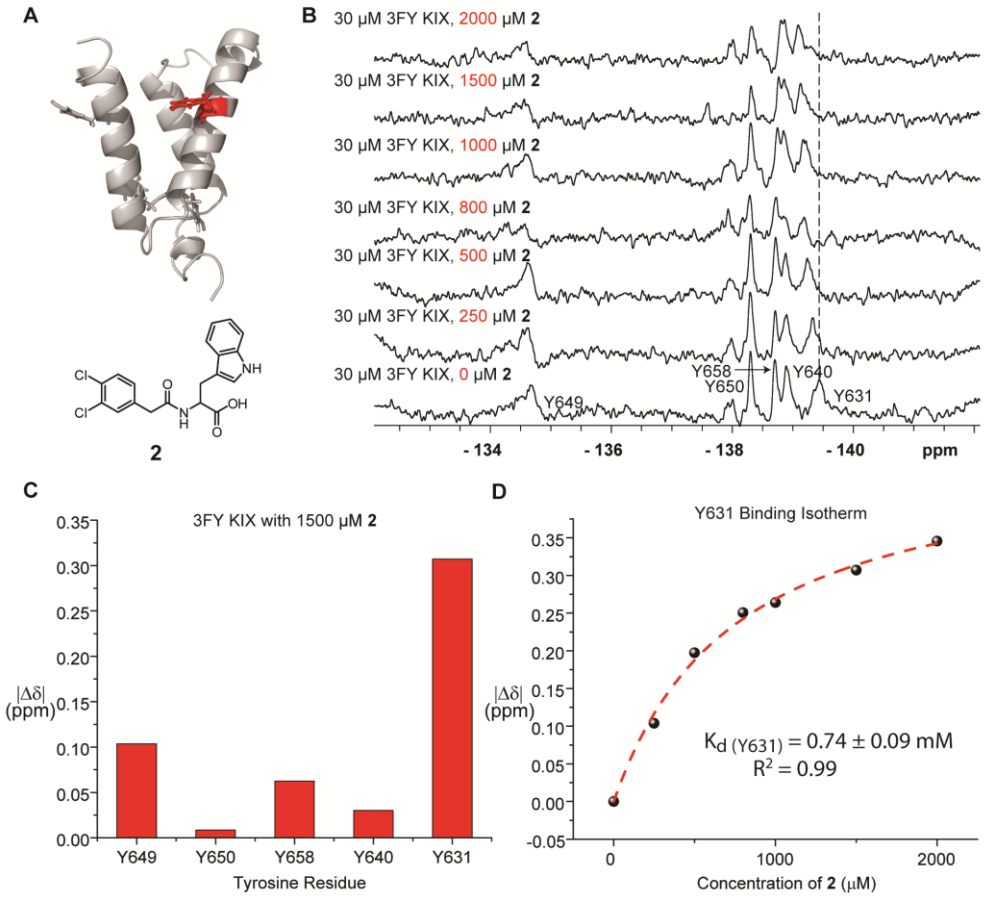


Figure 3.9 ^{19}F NMR spectral analysis of 3FY KIX in the presence of increasing concentrations of 2.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. B) Stacked spectra of 3FY KIX with increasing concentrations of **2** C) Absolute value of chemical shift perturbations for all 3FY KIX resonances at 1500 μM **2**. D) Binding isotherm of Y631 perturbation for the titration of **2**

Table 3.5 3FY-KIX ^{19}F NMR chemical shift perturbations at varying **2** concentrations

[**2**] (μM)	Chemical shift (ppm)	Y649	Y650	Y658	Y640	Y631
0	δ (ppm)	-134.6638	-138.3012	-138.7056	-138.8887	-139.437
250	|Δδ| (ppm)	0.0451	0.0048	0.0061	0.0099	0.1038
500	|Δδ| (ppm)	0.0503	0.0019	0.0110	0.0060	0.1974
800	|Δδ| (ppm)	0.0910	0.0120	0.0181	0.0250	0.2507
1000	|Δδ| (ppm)	0.0840	0.0014	0.0410	0.0428	0.2641
1500	|Δδ| (ppm)	0.1034	0.0085	0.0626	0.0301	0.3071
2000	|Δδ| (ppm)	0.0657	0.0186	0.1008	0.0231	0.3456

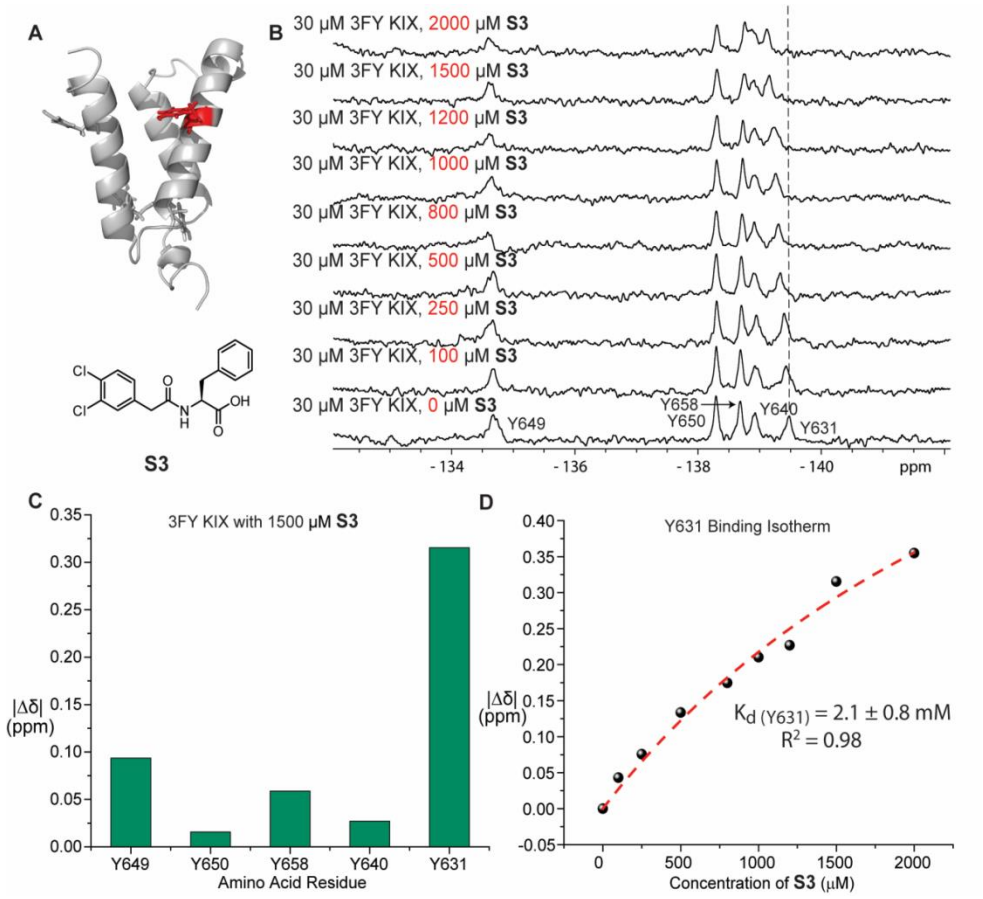


Figure 3.10 ^{19}F NMR spectral analysis of 3FY KIX in the presence of increasing concentrations of S3.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. B) Stacked spectra of 3FY KIX with increasing concentrations of **S3** C) Absolute value of chemical shift perturbations for all 3FY KIX resonances at 1500 μM **S3**. D) Binding isotherm of Y631 perturbation for the titration of **S3**

Table 3.6 3FY-KIX ^{19}F NMR chemical shift perturbations at varying S3 concentrations

[**S3**] (μM)	Chemical shift (ppm)	Y649	Y650	Y658	Y640	Y631
0	δ (ppm)	-134.6927	-138.2924	-138.6904	-138.9366	-139.4730
100	|Δδ| (ppm)	0.0111	0.0045	0.0009	0.0132	0.0431
250	|Δδ| (ppm)	0.0212	0.0114	0.0139	0.0094	0.0757
500	|Δδ| (ppm)	0.0170	0.0073	0.0138	0.0335	0.1336
800	|Δδ| (ppm)	0.0357	0.0053	0.0208	0.0270	0.1747
1000	|Δδ| (ppm)	0.0397	0.0101	0.0335	0.0081	0.2104
1200	|Δδ| (ppm)	0.0959	0.0180	0.0394	0.0140	0.2269
1500	|Δδ| (ppm)	0.0936	0.0156	0.0587	0.0269	0.3154
2000	|Δδ| (ppm)	0.0956	0.0157	0.0719	0.0554	0.3549

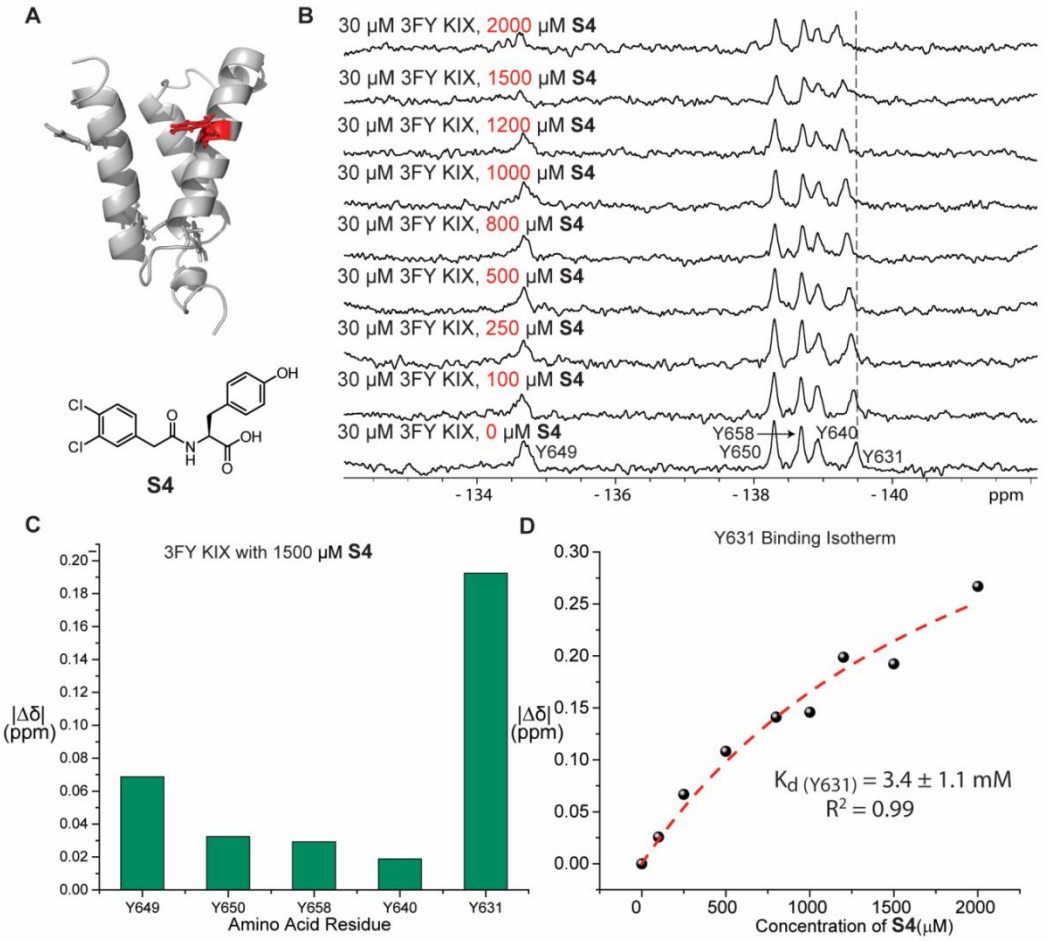


Figure 3.11 ^{19}F NMR spectral analysis of 3FY KIX in the presence of increasing concentrations of S4.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. B) Stacked spectra of 3FY KIX with increasing concentrations of **S4** C) Absolute value of chemical shift perturbations for all 3FY KIX resonances at 1500 μM **S4**. D) Binding isotherm of Y631 perturbation for the titration of **S4**

Table 3.7 3FY KIX ^{19}F NMR chemical shift perturbations at varying S4 concentrations

[**S4**] (μM)	Chemical shift (ppm)	Y649	Y650	Y658	Y640	Y631
0	δ (ppm)	-134.6927	-138.2924	-138.6904	-138.9366	-139.4730
100	|Δδ| (ppm)	0.0471	0.0008	0.0098	0.0125	0.0258
250	|Δδ| (ppm)	0.0167	0.0067	0.0083	0.0026	0.0668
500	|Δδ| (ppm)	0.0124	0.0128	0.0027	0.0067	0.1082
800	|Δδ| (ppm)	0.0042	0.0116	0.0146	0.0099	0.1411
1000	|Δδ| (ppm)	0.0147	0.0267	0.0192	0.0063	0.1458
1200	|Δδ| (ppm)	0.0325	0.0186	0.0224	0.0307	0.1987
1500	|Δδ| (ppm)	0.0688	0.0324	0.0292	0.0188	0.1924
2000	|Δδ| (ppm)	0.0739	0.0198	0.0331	0.0040	0.2669

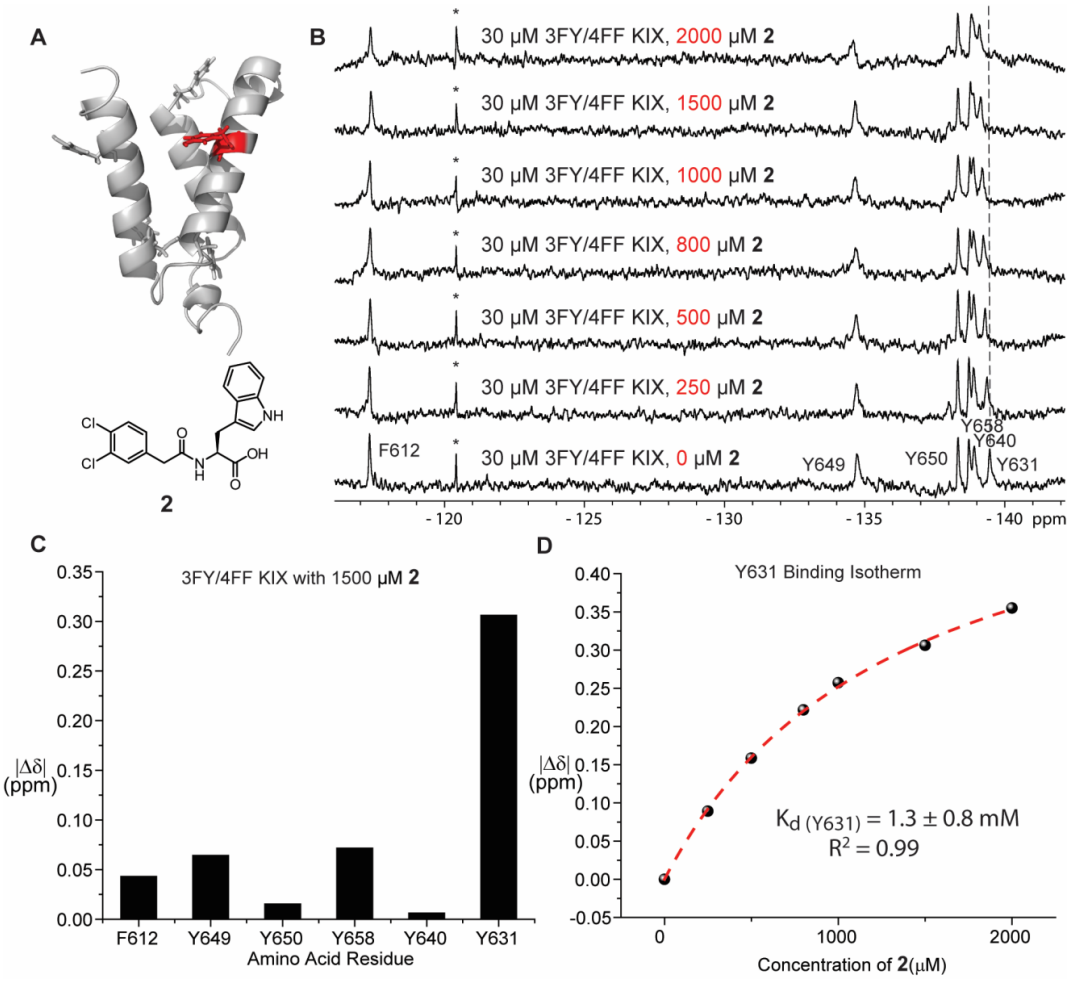


Figure 3.12 ^{19}F NMR spectral analysis of 3FY/4FF KIX in the presence of increasing concentrations of 2.

A) Ribbon diagram of KIX (PDB Code: 1KDX) with tyrosine side chains indicated as sticks. B) Stacked spectra of 3FY/4FF KIX with increasing concentrations of **2** * This resonance is an artifact that does not originate in the sample. C) Absolute value of chemical shift perturbations for all 3FY/4FF KIX resonances at 1500 μM **2**. D) Binding isotherm of Y631 perturbation for the titration of **2**.

Table 3.8 3FY/4FF-KIX ^{19}F NMR chemical shift perturbations at varying 2 concentrations

[2]	Chemical shift (ppm)	F612	Y649	Y650	Y658	Y640	Y631
0	δ (ppm)	-117.3300	-134.7212	-138.3184	-138.7131	-138.8843	-139.4560
250	|Δδ| (ppm)	0.0187	0.0361	0.0030	0.0100	0.0119	0.0892
500	|Δδ| (ppm)	0.0260	0.0278	0.0180	0.0351	0.0143	0.1586
800	|Δδ| (ppm)	0.0428	0.0352	0.0177	0.0396	0.0100	0.2218
1000	|Δδ| (ppm)	0.0362	0.0605	0.0146	0.0568	0.0056	0.2573
1500	|Δδ| (ppm)	0.0433	0.0643	0.0155	0.0718	0.0063	0.3063
2000	|Δδ| (ppm)	0.0376	0.0739	0.0153	0.0884		0.3551

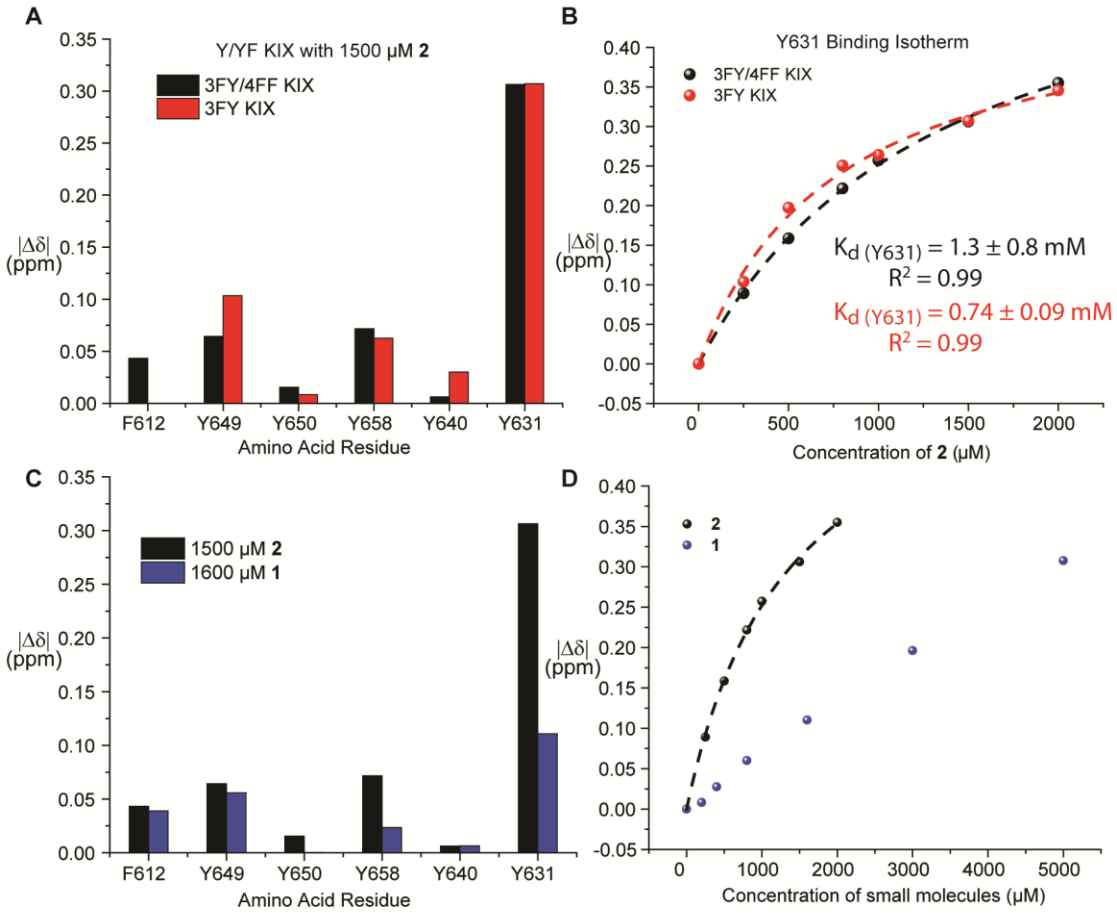


Figure 3.13 Comparison of binding data with 2.
A) Chemical shift perturbations of 2 with 3FY KIX and 3FY/4FF KIX demonstrating very similar perturbations. B) Binding isotherm of Y631 perturbations for the titrations of **2** against 3FY KIX and 3FY/4FF KIX. C) Chemical shift perturbations of **1** and **2** against 3FY/4FF KIX. Y631 is perturbed more significantly by **2** than it is by **1**. D) Binding isotherm of Y631 perturbations for the titrations of **1** and **2** against 3FY/4FF KIX.

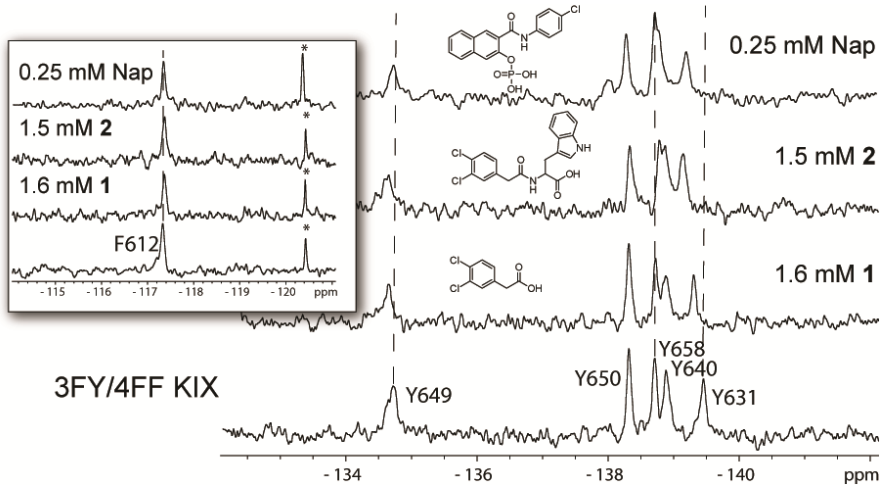


Figure 3.14 Small molecule binding to 3FY/4FF KIX.
Y631 shows significant perturbations while F612 (inset) shows no significant perturbations to chemical shift upon small molecule binding. The concentration 3FY/4FF KIX in the samples was 35 μM. Comparable ligand concentrations of **1** and **2** (1.6 mM and 1.5 mM, respectively) were used for chemical shift perturbation comparisons while 0.25 mM naphthol AS-E phosphate was used as a control.

Our PrOF NMR data led us to reevaluate the data that initially pointed towards the conclusion that **1** and its related analogs bound to KIX in the MLL site. We turned to the FTMap server to investigate possible binding sites in KIX.[202-205] As expected, the software identified the MLL and pKID binding sites with clusters of its probe solvent molecules. However, additional clusters were found distal to these binding sites yet still near Y631, the residue most perturbed in our PrOF NMR experiments (Figure 3.15B). Further, these clusters were situated in close proximity to amino acid residues perturbed by naphthol AS-E phosphate in previous ^{1}H-^{15}N HSQC experiments by Best et al. These clusters were also situated close to the residues perturbed by **1** in our previous ^{1}H-^{15}N HSQC validation studies. Clusters were consistently found in similar locations within the protein across multiple KIX PDB structures, lending more weight to this potential, previously unreported ligand binding site (1KDX, 2AGH, 2LXT, 2LXS, Figure 3.16).

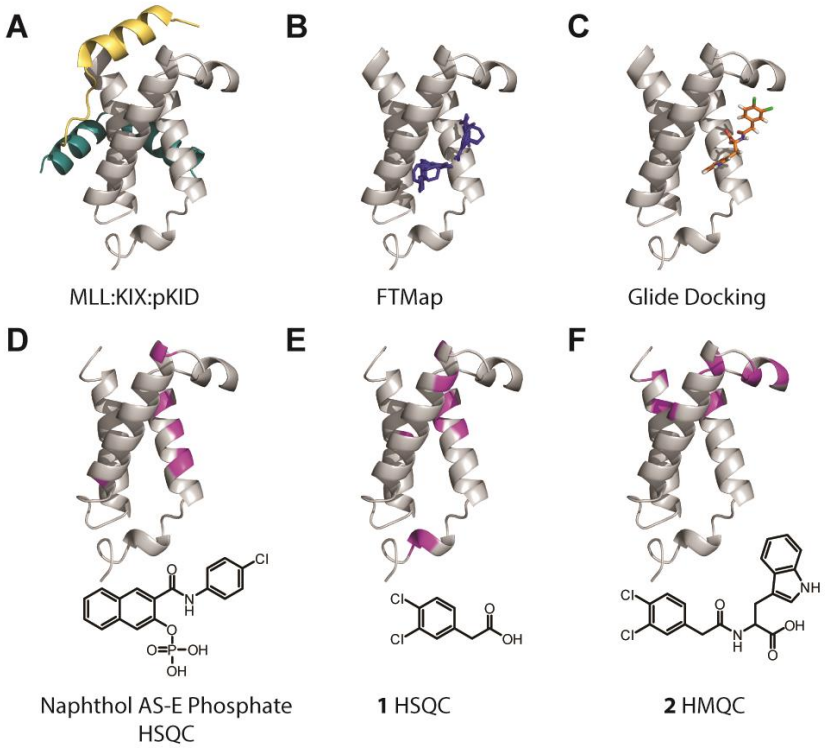


Figure 3.15 Structural representations of KIX highlighting binding sites mapped onto 1KDX.

A) Ternary complex of MLL, KIX, and pKID. B) Select FTMap results displaying probe molecule clusters in a third site in the protein (PDBID 1KDX). C) Glide docking pose of **2** using the Schrödinger Maestro Suite. D, E, F) Chemical shift perturbations from ^{1}H-^{15}N HSQC or HMQC experiments. Perturbations greater than one standard deviation are shown in the backbone.

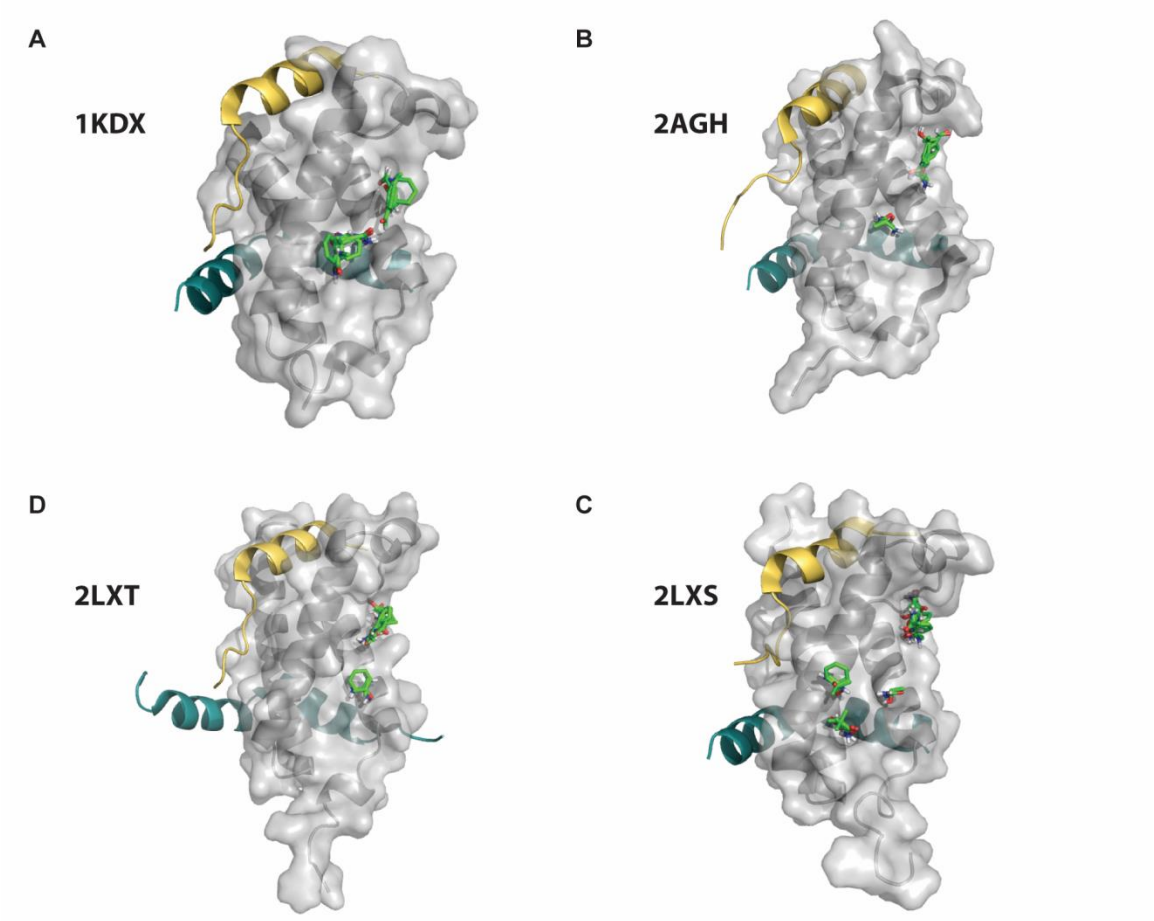


Figure 3.16 FTMap comparison of multiple KIX structures.
KIX ribbon structures are shown in gray with the native MLL peptide at the top, and the pKID peptide around the back. FTMap solvent probe clusters that do not occupy either the MLL or pKID site are shown in green, indicating the possibility of a new, previously unreported binding site.

As an additional followup, Glide docking studies in Schrödinger were performed with naphthol AS-E phosphate and **2** against the four previously listed KIX PDB structures. Grids large enough to encompass both the MLL binding site and this potentially new binding site were used to provide an unbiased comparative prediction. Using extra precision (XP) docking, two poses for each molecule were generated per PDB structure. Six out of eight poses for naphthol AS-E phosphate were found to be in the new binding site while two were predicted to be in the MLL site. For **2**, all eight

poses were found in the new binding site consistent with the clusters predicted by FTMap (Figure 3.15C, Figure 3.17).

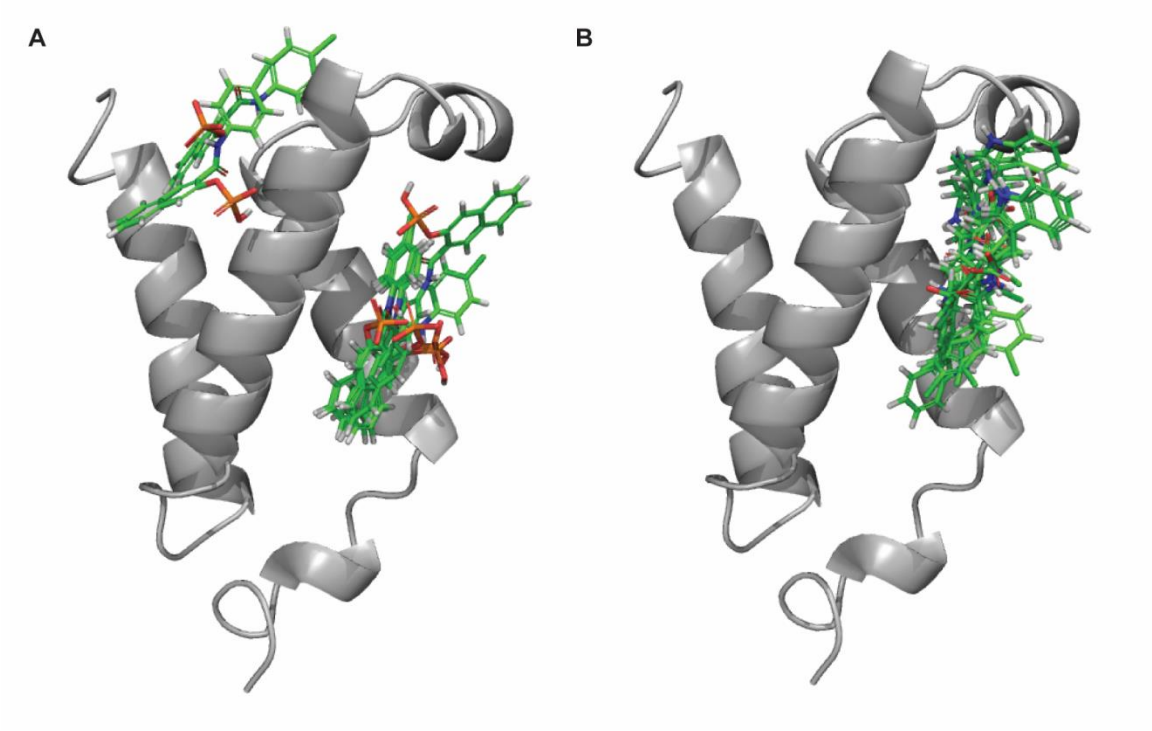


Figure 3.17 Glide docking pose comparisons of small molecule ligands against four different KIX PDBs (1KDX, 2AGH, 2LXT, 2LXS).
Two poses per PDB are presented. A) Naphthol AS-E phosphate binding poses predicted by Glide docking. Six poses place naphthol AS-E phosphate in the new binding site while the remaining two are placed in the MLL site. B) Binding poses of **2** predicted by Glide docking. All eight predicted poses are in the new binding site, and no poses are predicted for the MLL site.

Reevaluation of the HSQC NMR data from Best et al. (naphthol AS-E phosphate)[36] and Gee et al. (**1**)[193] revealed the possibility of a new ligand binding site, consistent with the most significant perturbations in the amide backbones being found on the α1 helix. ^{1}H-^{15}N HMQC NMR data on **2** resulted in similar perturbations to both naphthol AS-E phosphate and **1**. (Figure 3.15D, E, F, Figure 3.18)

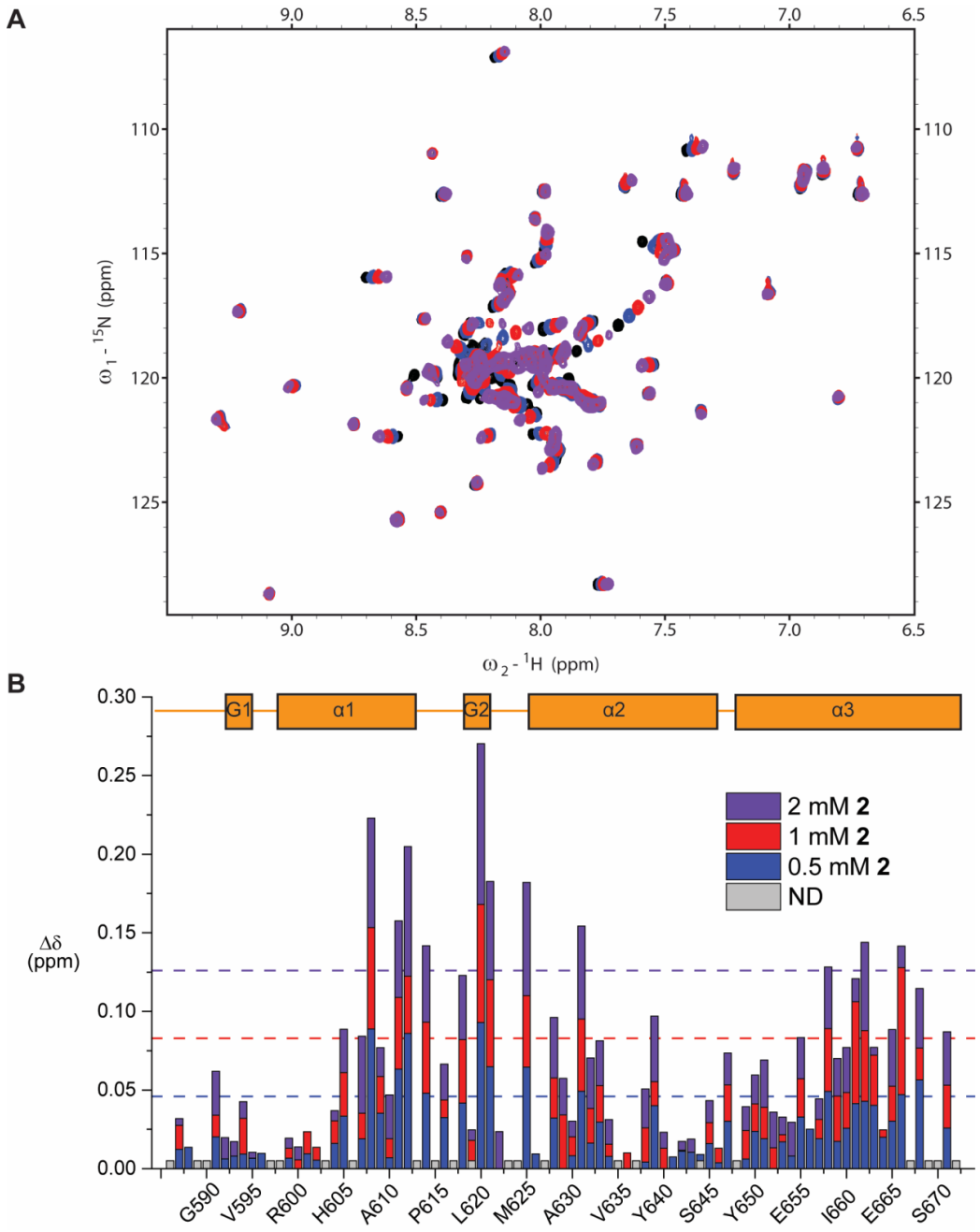


Figure 3.18 ^{1}H-^{15}N HSQC of 2 binding to KIX.
A) Spectra of ^{15}N KIX alone, ^{15}N KIX + 0.5 mM **2**, ^{15}N KIX + 1 mM **2**, and ^{15}N KIX + 2 mM **2** B) Chemical shift perturbation mapping of KIX residues with 0.5 mM, 1 mM and 2 mM **2**. Residues that were not detected or not assigned are designated as ND. Dashed lines represent the average chemical shift perturbation + one standard deviation at the corresponding concentration. Assignments previously obtained by Majmudar et al.[42] In the secondary structure box schematic, G1 and G2 refer to the two 3_{10}-helices, while α1, α2, and α3, refer to the three alpha helices comprising the structure

Notably, in contrast to **1**, vectors for the chemical shift perturbations in the presence of increasing concentrations of **2** were consistently linear, indicative of 1:1 binding (Figure S15). Surprisingly, one of the most significantly perturbed amide backbones was that of F612. This apparent contradiction could be the result of a number of different factors. The 4FF may be sensitive enough to detect stronger binders like the native MLL peptide but not sensitive enough to detect weaker binders that are farther away. As an additional distinction HMQC experiments reports on changes in the chemical environment of the amide backbone while PrOF NMR reports on changes in the chemical environment of the fluorine on the side chain, which may be less sensitive to local secondary structure effects. It is also possible that the chemical environment of this solvent-exposed 4FF has not changed sufficiently, relative to the change in chemical environment near the backbone amide, due to either direct or distant binding effects. Additional experiments are ongoing to test the sensitivity of other fluorinated phenylalanine variants.

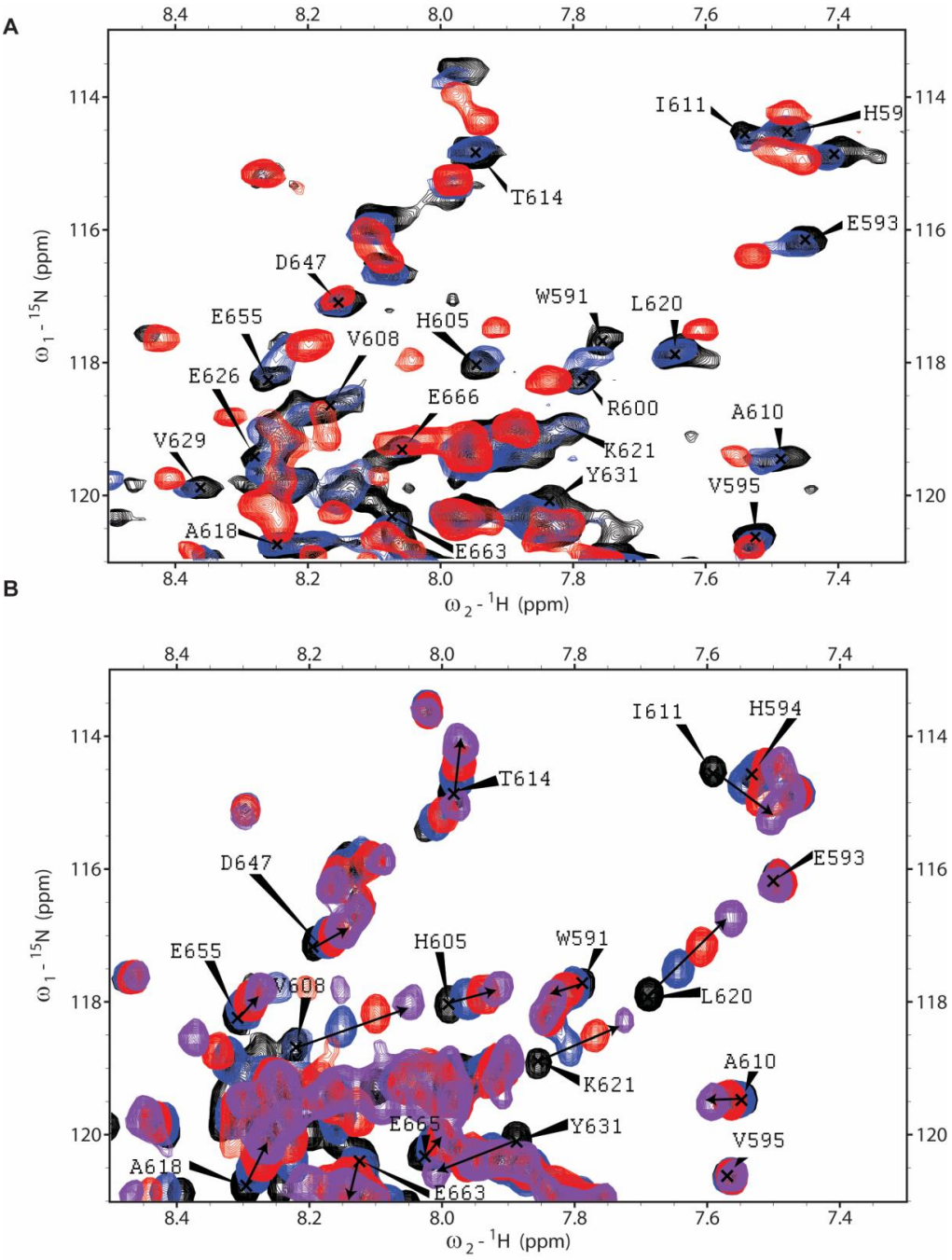


Figure 3.19 Comparison of ^{1}H-^{15}N HSQC/HMQC data from KIX.
A) Select region of ^{1}H-^{15}N HSQC NMR spectrum with **1**. Native protein, 2 mM **1** and 5 mM **2** are shown. B) Select region of ^{1}H-^{15}N HMQC NMR spectrum with **2**. Distinctly, the chemical shift perturbations in the presence of **2** consistently exhibit a linear change in chemical shift relative to the parent compound **1**. Native protein, 0.5 mM **2**, 1 mM **2**, and 2 mM **2** are shown. Assignments previously obtained by Majmudar et al.[42]

3.3 **Conclusion**

This dual-labeling strategy augments previously reported PrOF NMR methods for studying ligand-protein interactions. Dual-labeling provides the ability to insert additional probes into the protein of interest and report on potentially new ligand binding modes. Using this novel dual-labeling strategy in conjunction with computational predictions and HSQC NMR data, our data supports a new small-molecule binding site in KIX, distal to the two native transcription factor binding sites. Such a new binding site provides an opportunity to positively or negatively allosterically regulate KIX PPIs, such as with naphthol AS-E phosphate, without competing with endogenous transcription factors. Identifying ligand binding modes in this site will be the subject of future studies. Although this method produces lower protein expression yields making it less ideal for full screens, dual-labeling a protein can serve as a useful tool and extension of PrOF NMR for further studying ligand binding and characterizing their binding modes.

3.4 **Future Directions**

There are still a number of directions that these projects could go. The most immediate follow-up would be to further study the discrepancy between the results from **2** binding to KIX via HSQC and PrOF NMR. The sensitivity of different fluorinated phenylalanines could be assessed to see if alternate substitutions would provide more insight into the ligand binding. Another round of experiments related to further studying the binding site would be to utilize fluorescence polarization. While **2** is not a potent enough ligand to compete with the native peptide sequences, it is possible that its binding could still allosterically regulate the binding of MLL, pKID, or both. One could first saturate the protein with **2**, then assess the affinity of either native peptide individually and together to see if their relative affinities are affected by the presence of **2**. One might expect that if **2** is stabilizing the fold of the protein through binding that the peptides may actually bind with a tighter affinity.

Additionally, more direct approaches could also be taken. With the appropriate mutations in the protein sequence, one could perform a disulfide tethering experiments or a photocrosslinking experiment to form a covalent bond between the protein and the

ligand.[8] Once bound, the protein could be digested, and analyzed to identify where the crosslink was formed. An alternative strategy would be to use a transfer-nuclear overhauser effect NMR experiment in which the ligand is irradiated and the protein signals are observed to identify which residues receive a transfer of energy. This type of experiment would help to identify which amino acids are in close proximity to the bound small molecule as a complementary approach to the HSQC experiments.

The clearest method to identify exactly where the molecules are binding in the protein would be to obtain x-ray crystal structures. However, KIX has been notoriously difficult to crystallize and has only been crystallized with a covalent ligand attached. In conjunction with the crosslinking idea, one could attempt crystallization experiments once the molecule is crosslinked to see if the covalent attachment is enough to help stabilize the protein. Crosslinking to 3FY KIX instead of unlabeled KIX would also provide additional stabilization as demonstrated by the increased T_m of 3FY KIX.

On the protein labeling side, it would still be interesting to work on other dual-labeling strategies with other combinations of amino acids. It could also be interesting to apply a dual labeling strategy to the bromodomains that are already being expressed in the lab. The bromodomains tend to express more efficiently. One of the main drawbacks to the current dual-labeling process is the need to use DL39 cells. We can already express 5-fluoro-tryptophan labeled proteins in regular BL21 cells, so if it were possible to express proteins with other fluorinated amino acids in a non-auxotrophic cell line, it might make the process significantly more efficient.

3.5 **Materials and Methods**

Materials

Unless otherwise noted, all reagents were obtained from commercial vendors and used without further purification. For expression of proteins in *E. coli*, Lennox L. (LB) broth, unlabeled amino acids, uracil, thiamine-HCl, nicotinic acid, biotin, ampicillin, chloramphenicol, IPTG, PMSF, and Triton X-100 were purchased from RPI corp. Nucleotide bases and 3-fluorotyrosine (3FY) were purchased from Alfa Aesar. Lysozyme was purchased from Gold Biotech, HEPES was purchased from Fischer Scientific, and

[^{15}N] ammonium chloride was obtained from Cambridge Isotope Laboratories. ^{1}H, ^{13}C, and ^{19}F spectra were recorded at 500 MHz, 125 MHz, and 470 MHz respectively. MLL and pKID peptide sequences, synthesis, and characterization were reported by Pomerantz et al.[41]

Methods:

Unlabeled (UL), 3FY, 4FF, and 3FY/4FF-Labeled Protein Expression and Purification of His$_6$-KIX (586-672)

Expression

Protein expression using the same plasmid DNA for KIX was performed as previously described[41] with slight modifications. The protein expression plasmid was transformed into either the *E. coli* BL21(DE3) strain co-transformed with pRARE (Novagen) (unlabeled protein) or DL39(DE3) cells with the plasmid pRARE (3FY-, 4FF-, or 3FY/4FF-labeled proteins) and plated on an agar plate containing ampicillin (100 mg/L) and chloramphenicol (35 mg/L). 5 mL primary cultures of DL39 or BL21 cells transformed with the KIX expression plasmid supplemented with ampicillin (100 mg/L) and chloramphenicol (34 mg/L) were grown overnight (~12-14 h, 25 °C, 250 rpm) in an incubator shaker (New Brunswick Scientific Excella E25). Primary cultures were added to fresh LB media (1 mL primary/50 mL media) to begin secondary culture growth. Secondary cultures were grown at (37 °C, 250 rpm) until they reached an OD_{600} of approximately 0.8 as evaluated by a Varian Cary 50 Bio UV/Vis spectrophotometer. The culture was centrifuged (8000 g, 10 min, 4 °C) to form a solid pellet of bacteria. The LB media was decanted, and the pellet was resuspended in defined media[187] containing the designated amount of aromatic or fluorinated aromatic amino acids (Table S1) at the original volume of LB media.. A recovery time of 30-120 min at 20 °C, 250 rpm was allowed prior to the addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) to induce protein overexpression overnight (15-20 h) Cultures were spun down (6000 g, 20 min, 4 °C). Media was decanted, and the pellet was stored at -20 °C.

Optimization of 4FF and 3FY/4FF KIX Expression Conditions

A range of conditions were varied to determine optimal expression conditions, including amount of 4FF, phenylalanine, and recovery time. Results are summarized in the table below. Yields were not recorded for many early experiments since the first goal was to optimize label incorporation conditions. Initial results indicated that additional phenylalanine should be kept under 40 μM. Follow-up studies showed little to no improvement in label incorporation with higher amounts of 4FF. The next round of experiments determined that a 30 minute recovery time and 10 μM provided improved label incorporation. Longer recovery times were also explored and slightly lower amounts of phenylalanine were explored, determining that between 7.5 and 10 μM F and a 30 – 40 min. recovery time yielded the best results. High labeling of the dual-labeled protein is also achievable by leaving out phenylalanine.

Purification via Fast Protein Liquid Chromatography (FPLC)

Following cell lysis (lysis buffer, lysozyme, β-mercaptoethanol, protease inhibitor phenylmethanesulfonylfluoride (PMSF, 2 mM), and sonication), proteins were purified on a GE Äktapurifier FPLC equipped with a Superloop for large volume injections. A column containing Ni-NTA beads (Qiagen) was used for affinity purification of the His-tagged protein. An elution gradient (30-400 mM) imidazole was used to initially bind and subsequently elute the His-tagged-protein from the Ni beads. Protein was then buffer exchanged into 50 mM HEPES and 100 mM NaCl, pH 7.2 using a HiPrep 26/10 Desalting column (GE Healthcare, Life Sciences) and stored at -20 °C or 4°C. During protein purification multiple aliquots of the protease inhibitor PMSF were added due to the instability of PMSF in aqueous solution.

UPLC/MS Conditions for Protein Characterization

For UPLC/MS analysis of proteins, a Waters Acquity UPLC coupled to a Waters Synapt G2 HDMS quadrupole orthogonal acceleration time of flight mass spectrometer was used (Waters Corp., Milford, MA USA). A Waters Acquity UPLC Protein BEH C_4 2.1 mm x 100 mm column (1.7 μm diameter particles) at 35°C was used for the following 15 min linear gradient separation at a flow rate of 0.400 mL/min using A: water containing 0.1% formic acid and B: acetonitrile containing 0.1% formic acid: 3% B, 0 min to 3 min; 3% B to 97% B, 3 min to 9 min; 97% B, 9 min to 11 min; 97% B to 3% B,

11 min to 13 min; 3% B 13 min to 15 min. Mass spectra were collected in profile mode over the range *m/z* 300-2500 every 0.1 s during the chromatographic separation. MS parameters in positive electrospray ionization mode were as follows: capillary, 0.3 kV; sampling cone, 35.0 V; extraction cone, 4.0 V; desolvation gas flow, 800 L/h; source temperature, 100 °C; desolvation temperature, 350 °C; cone gas flow, 20 L/h; trap CE, off. Lockspray (on-the-fly mass calibration) configuration consisted of infusion of a 5 μg/mL solution of leucine-enkephalin and acquisition of one mass spectrum (0.2s scan, *m/z* 50-1200) every 10s. Three lockspray *m/z* measurements of protonated (positive ionization mode) leucine-enkephalin were averaged and used to apply a mass correction to measured *m/z* values during the course of the analysis.

Circular Dichroism

CD spectra were acquired using a peltier equipped temperature controlled Jasco J-815 spectropolarimeter. Protein samples were prepared in a 10 mM sodium phosphate, 100 mM sodium chloride buffer at pH 7.2. Unlabeled KIX was prepared at 10.5 μM while 3FY and 3FY/4FF KIX were prepared at 11.5 μM. Molar ellipticity was monitored from 200-260 nm. Samples were collected at 25 °C using a 1 mm cuvette path-length. Reported data is the average of 5 acquired spectra. Thermal stability of unlabeled and fluorine-labeled KIX was measured by following the change in ellipticity at 222 nm from 20 °C to 95 °C. For the thermal melt experiments, the concentration of 3FY/4FF KIX was increased to 35 μM. The T_m was calculated using a sigmoidal fit. Data analysis was performed using Origin 2016. Molar ellipticity was calculated using the following equation.

$$\Theta_{222} = \Psi/(1000nlc) \quad (1)$$

In equation 1, Ψ is the CD signal in degrees, *n* is the number of amides in the protein construct, *l* is the path lenth in centimeters, and *c* is the concentration of the protein sample in decimoles per cubic centimeter.

Fluorescence Polarization

Fluorescence polarization experiments were performed on a Tecan Infinte F500. A FITC-MLL peptide[41] was used to compare the binding activity of the three different protein constructs. UL and 3FY/4FF KIX were prepared at a starting concentration of 40

μM while 3FY KIX was prepared at a starting concentration of 35 μM. Samples also contained 85 nM tracer (FITC-MLL), 0.01% Triton X-100, with a total DMSO concentration of 0.5%. Excitation and emission wavelengths were 485 nm and 535 nm respectively. Samples were incubated for 30 min at room temperature prior to scanning. Each data point in Figure S2 C is the average of three independent experiments with the standard deviations listed as the error. Data analysis was performed with GraphPad Prism. K_d values were determined as follows

$$y = c + (b - c) * \frac{\sqrt{(K_d + a + x)^2 - 4ax}}{2a} \quad (2)$$

In the above equation, y corresponds to the observed anisotropy, c and b and the minimum and maximum observed entropies, respectively, K_d is the dissociation constant, and a and x are the concentrations of fluorescent peptide and KIX, respectively,

1D ^{19}F NMR.

Conditions and Parameters Used. All NMR spectra were obtained on a Bruker Avance III 500 MHz instrument equipped with a 5 mm Prodigy TCI Cryoprobe (^{19}F S/N 2100:1). Samples containing 30-50 μM 3FY, 4FF, or 3FY/4FF KIX were prepared in NMR buffer as noted above with 5% D_2O and 50 μM trifluoroacetic acid (-76.5 ppm) as a reference. T_1 relaxation times were calculated using an inversion recovery experiment. Optimized parameters for PrOF NMR experiments included a 0.6 s delay time, a 0.1-0.3 s acquisition time (AQ), and 400-800 scans. Spectral width and offset were 10 ppm and -136 ppm respectively for 3FY KIX. Spectral width and offset were 10 ppm and -115 ppm respectively for 4FF KIX. Both of the aforementioned parameter sets were used when analyzing the 3FY/4FF KIX in addition to an experiment with a spectral width of 26 ppm and offset of -128 ppm for a full spectrum containing the 3FY and 4FF resonances. A line-broadening of 15-20 Hz was applied to all protein spectra upon processing. Spectra were acquired at 294 K)

PrOF NMR K_d values are determined using the following equation:

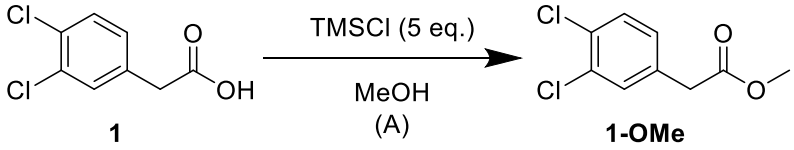
$$S = D * \frac{(K_d + L + P) - \sqrt{(K_d + L + P)^2 - 4PL}}{2P} \quad (3)$$

In equation 2, S corresponds to the observed change in chemical shift, D is the maximum change in chemical shift, Kd is the dissociation constant of the ligand, L is the ligand concentration, and P is the protein concentration.

2D ^{1}H-^{15}N HSQC NMR Experiments

Uniformly ^{15}N labeled KIX protein was expressed and purified using M9 minimal media supplemented with [^{15}N] ammonium chloride as described previously.[42] A 30-35 μM solution of ^{15}N-labeled KIX was prepared in a 9:1 H_2O:D_2O 10 mM phosphate buffer with 100 mM NaCl and 2.5% DMSO at pH 7.2. HSQC experiments were recorded at 27 °C on a Bruker Ascend 850 MHz NMR spectrometer equipped with a 5 mM TCI Cryo probe. Spectra were collected with 0, 0.5, 1, and 2 mM **2**. Data was processed using NMRPipe and analyzed using Sparky (UCSF). Chemical shift changes for individual resonances were quantified using the equation $((0.2*\Delta\delta^{15}N)^2 + (\Delta\delta^{1}H)^2)^{0.5}$.

Small Molecule Synthesis

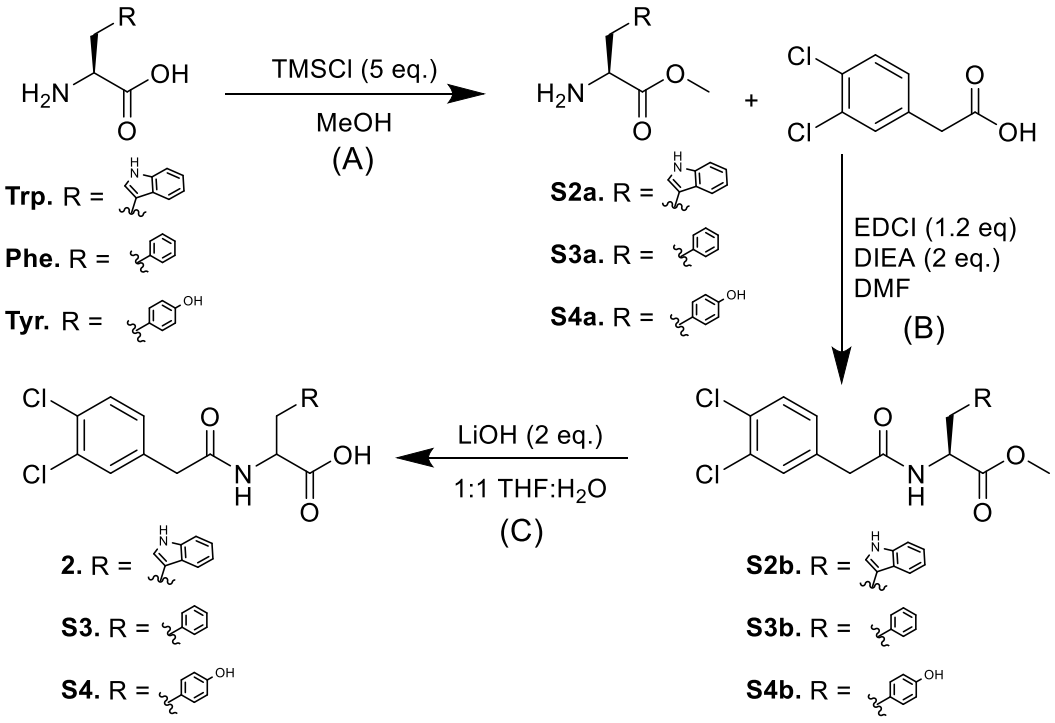


Scheme 3.1 Methyl esterification of carboxylic acids

***General Procedure A*: *Methyl esterification of carboxylic acids*:** One of the compounds **1** or amino acids tryptophan, tyrosine, or phenylalanine, was dissolved in 5-10 mL MeOH. TMSCl (5 eq) was added, and the reaction was stirred at room temperature overnight. TLC indicated the reaction was complete. The solvent was evaporated *in vacuo*, and material was collected to carry on to the next step without further purification.

methyl 2-(3,4-dichlorophenyl)acetate (1-OMe)

3,4-Dichlorophenylacetic acid (**1**) was converted to its methyl ester according to general procedure A on a 0.512 mmol scale. The purification of this compound deviated from the general procedure since it was the final compound. After removing the MeOH solvent, the material was dissolved in EtOAc, and the organic phase was washed three times with a $NaHCO_3$ solution, and then washed with brine. The organic layer was removed, dried over $MgSO_4$, filtered to remove the drying agent, and concentrated *in vacuo*, providing **1-OMe** as a white solid (80 mg/0.365 mmol, 71%): ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.39 (d, *J* = 8.1 Hz, 1 H), 7.38 (d, *J* = 2.0 Hz, 1 H), 7.11 (dd, *J* = 8.1, 2.0 Hz, 1 H), 3.70 (s, 3 H), 3.57 (s, 2 H); ^{13}C NMR (125 MHz, $CDCl_3$): δ 171.14, 134.10, 132.64, 131.51, 131.41, 130.58, 128.88, 52.40, 40.21;



Scheme 3.2 Analog synthesis for ligand development

General Procedure B*: *Amide bond formation Compound **1** and **S2a**, **S3a**, or **S4a** (1 eq) were dissolved in DMF with 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI) (1.2 eq) and N,N-diisopropylethylamine (DIEA) (2 eq). Reactions were stirred at room temperature overnight. TLC indicated the completion of the reaction. The reaction mixture was then diluted with EtOAc and washed with water and brine. The organic layer was collected, dried over $MgSO_4$, and the solvent was evaporated *in vacuo*. The product was then purified by column chromatography and isolated as a solid.

General Procedure C: Hydrolysis of methyl ester to carboxylic acid. Compound **S2b**, **S3b**, or **S4b** (1 eq) were dissolved in a 1:1 THF:H_2O solution with LiOH (2 eq) and stirred for approximately 6 hours. TLC indicated the completion of the reaction. The final product was isolated using an acid/base extraction to separate starting materials and unreacted methyl esters from the free carboxylic acid. The final organic layer was dried over $MgSO_4$, filtered, and concentrated *in vacuo* to isolate a solid. It is important to note that the stereochemistry of the final product (**2, S3, S4**) is not indicated in the reaction scheme because it is believed that racemization may occur during hydrolysis.

Tryptophan Analogs

Methyl Tryptophanate (S2a)

Tryptophan (**Trp**) was converted to its methyl ester according to general procedure A on a 1.47 mmol scale in 10 mL of MeOH. Removal of solvent yielded **S2a** as a red-purple solid and was used without further purification. ^{1}H NMR (500 MHz, MeOD): δ 7.54 (d, *J* = 7.9 Hz, 1 H), 7.39 (d, *J* = 8.1 Hz, 1 H), 7.19 (s, 1 H), 7.15 (t, *J* = 7.6 Hz, 1 H), 7.07 (t, 7.6 Hz, 1 H), 4.32 (dd, *J* = 7.6, 5.5 Hz, 1 H), 3.80 (s, 3 H), 3.46 (dd, *J* = 15.1, 5.4, 1 H), 3.35 (dd, *J* = 15.1, 7.6 Hz, 1 H); ^{13}C NMR (125 MHz, MeOD): δ 170.86, 138.36 (low intensity), 128.17, 125.53, 123.00, 120.36, 118.79, 112.70, 107.49, 54.64, 53 65, 27.62; Hi-Res MS calc'd *m/z*, 218.10553, found M+H [219.11366]

Methyl (2-(3,4-Dichlorophenyl)acetyl)tryptophanate (S2b)

Compound **S2a** was coupled to **1** according to general procedure B on a 1.18 mmol scale. Following the removal of solvent *in vacuo*, the material was purified by column chromatography using a 3:2 EtOAc:Hexanes mobile phase. The final product (**S2b**) was isolated as a yellow solid (199 mg/0.491 mmol, 41.5% yield). ^{1}H NMR (500 MHz, MeOD): δ 7.47 (d, *J* = 7.9 Hz, 1 H), 7.30-7.34 (m, 3 H), 7.09 (t, *J* = 7.6 Hz, 1 H), 6.95-7.01 (m, 3H), 4.76 (dd, *j* = 8.3, 5.3 Hz 1 H), 3.68 (s, 3 H), 3.45 (s, 2 H), 3.3 (m, 1 H)*, 3.16 (dd, *J* = 14.7, 8.3 Hz, 1 H) *Signal obscured due to solvent overlap; ^{13}C NMR (125 MHz, MeOD): δ 173.82, 172.70, 138.02, 137.35, 133.08, 132.21, 131.69, 131.38, 129.98, 128.69, 124.36, 122.47, 119.87, 119.08, 112.36, 110.62, 54.98, 52.73, 42.16, 28.35; Hi-Res MS calc'd *m/z*, 404.06945 found M+Na [427.06065]

(2-(3,4-Dichlorophenyl)acetyl)tryptophan (2)

Compound **S2b** was hydrolyzed according to general procedure C on a 0.10 mmol scale. The final compound **2** was isolated as a yellow oil (26 mg/, 66.5% yield). ^{1}H NMR (500 MHz, MeOD): δ 7.51 (d, *J* = 8.0 Hz, 1 H), 7.33 (d, *J* = 8.1 Hz, 1 H), 7.29 (s, 1 H), 7.28 (d, *J* = 8.1 Hz, 1 H), 7.09 (ddd, *J* = 7.9, 7.2, 1.0 Hz, 1 H), 7.03 (s, 1 H), 6.98 (ddd, *J* = 7.9, 7.2, 1.0 Hz, 1 H), 6.92 (dd, *J* = 8.2, 1.9 Hz, 1 H), 4.75 (dd, *J* = 8.3, 4.8 Hz, 1 H), 3.43 (s, 2 H), 3.37 (dd, *J* = 14.8, 4.7 Hz, 1 H), 3.17 (dd, *J* = 14.7, 8.3 Hz, 1 H); ^{13}C NMR (125 MHz, MeOD): δ 175.64, 172.54, 137.97, 137.36, 133.05, 132.27, 131.64, 131.37, 130.01, 128.93, 124.33, 122.36, 119.81, 119.23, 112.29, 111.04, 55.15, 42.35, 28.42; Hi-Res MS calc'd *m/z*, 390.05380 found M-H [389.04570]

Phenylalanine Analogs

Methyl Phenylalaninate (S3a)

Phenylalanine (**Phe**) was converted to its methyl ester according to general procedure A on a 1.82 mmol scale in 10 mL of MeOH. Removal of solvent yielded **S3a** as a white solid and used without further purification. ^{1}H NMR (500 MHz, MeOD): δ 7.38 (t, *J* = 7.3 Hz, 2 H), 7.33 (t, *J* = 7.3 Hz, 1 H), 7.26 (d, *J* = 7.3 Hz, 1 H), 4.32 (dd, *J* = 6.2, 7.3 Hz, 1 H), 3.80 (s, 3 H), 3.27 (dd, *J* = 14.4, 6.2 Hz, 1 H), 3.17 (dd, *J* = 14.4, 6.2 Hz, 1H); ^{13}C NMR (125 MHz, MeOD): δ 170.47, 135-27, 130.40, 130.20, 129.01, 55.20, 53.60, 37.42; Hi-Res MS calc'd *m/z*, 179.19463 found M+H [180.10942]

Methyl (2-(3,4-Dichlorophenyl)acetyl)phenylalaninate (S3b)

Compound **S3a** was coupled to **1** according to general procedure B on a 0.858 mmol scale. Following the removal of solvent *in vacuo*, the material was purified by column chromatography using a 2:1 EtOAc:Hexanes mobile phase. The final product (**S3b**) was isolated as a white solid (116 mg/0.317 mmol, 37.0 % yield). ^{1}H NMR (500 MHz, MeOD): δ 7.48 (d, *J* = 8.3 Hz, 1 H), 7.35 (d, *J* = 2.0 Hz, 1 H), 7.19-7.24 (m, 3 H), 7.13 (dd, *J* = 8.0, 1.7 Hz, 2 H), 7.06 (dd, *J* = 8.4, 2.0 Hz, 1 H), 4.68 (dd, *J* = 9.4, 5.3 Hz, 1 H), 3.71 (s, 3 H), 3.45 (s, 2 H), 3.17 (dd, *J* = 13.9, 5.2 Hz, 1 H), 2.94 (dd, *J* = 13.9, 9.6 Hz, 1 H); ^{13}C NMR (125 MHz, MeOD): δ 173.35, 172.69, 138.02, 137.44, 133.12, 132.16, 131.76, 131.45, 130.15, 130.05, 129.48, 127.89, 55.18, 42.77, 42.24, 38.23; Hi-Res MS calc'd *m/z*, 365.05855 found M+Na [388.05118]

(2-(3,4-Dichlorophenyl)acetyl)phenylalanine (S3)

Compound **S3b** was hydrolyzed according to general procedure C on a 0.082 mmol scale. The final compound **S3** was isolated as a white solid, though a small contamination does not allow for accurate yield calculations. ^{1}H NMR (500 MHz, MeOD): δ 7.37 (d, *J* = 8.3 Hz, 1 H), 7.34 (d, *J* = 2.0 Hz, 1 H), 7.18-7.24 (m, 3 H), 7.13 (dd, *J* = 8.0, 1.1 Hz, 2 H), 7.04 (dd, *J* = 8.4, 2.0 Hz, 1 H), 4.66 (dd, *J* = 9.4, 5.0 Hz, 1 H), 3.45 (d, *J* = 2.0 Hz, 2 H), 3.22 (dd, *J* = 14.0, 4.9 Hz, 1 H), 2.93 (dd, *J* = 14.0, 9.4 Hz, 1 H); ^{13}C NMR (125 MHz, MeOD): δ 174.80 (low intensity), 172.55, 138.37, 137.53, 133.11, 132.21, 131.73 (low intensity), 131.42, 130.22, 130.07, 129.39, 127.74, 55.14, 42.38, 38.35; Hi-Res MS calc'd *m/z*, 351.04290, found M-H [350.02350]

Tyrosine Analogs

Methyl Tyrosinate (S4a)

Tyrosine (**Tyr**) was converted to its methyl ester according to general procedure A on a 1.66 mmol scale in 10 mL of MeOH. Removal of solvent yielded **S4a** as a white solid (126.4 mg/0.648 mmol, 39% yield). ^{1}H NMR (500 MHz, $CDCl_3$/MeOD): δ 6.96 (d, *J* = 8.3 Hz, 2 H), 6.73 (d, *J* = 8.3 Hz, 2 H), 3.69 (s, 3 H), 3.65 (dd, *J* = 7.6, 5.4 Hz, 1 H), 2.97 (dd, *J* = 13.7, 5.2 Hz, 1 H), 2.78 (dd, *J* = 13.7, 7.6 Hz, 1 H); ^{13}C NMR (125 MHz, $CDCl_3$/MeOD): δ 175.28, 155.79, 130.37, 127.61, 115.65, 55.56, 52.18, 39.78; Hi-Res MS calc'd *m/z*, 195.08954 found M+H [196.09711]

Methyl (2-(3,4-Dichlorophenyl)acetyl)tyrosinate (S4b)

Compound **S4a** was coupled to **1** according to general procedure B on a 0.647 mmol scale. Following the removal of solvent *in vacuo*, the material was purified by column chromatography using a 79:1 DCM:MeOH mobile phase. The final product (**S4b**) was isolated as a white solid (44.9 mg/0.117 mmol, 18.0 % yield). ^{1}H NMR (500 MHz, MeOD): δ 7.39 (d, *J* = 8.3, 1 H), 7.37 (d, *J* = 2.0 Hz, 1 H), 7.03 (dd, *J* = 8.2, 2.0 Hz, 1 H), 6.95 (d, *J* = 8.5 Hz, 2 H), 6.67 (d, *J* = 8.6 Hz, 2 H), 4.62 (dd, *J* = 9.2, 5.4 Hz, 1 H), 3.70 (s, 3 H), 3.47 (s, 2 H), 3.07 (dd, *J* = 14.1, 5.4 Hz, 1 H), 2.85 (dd, *J* = 14.1, 9.2 Hz, 1 H) ^{13}C NMR (125 MHz, MeOD): δ 173.50, 172.69, 157.47, 137.47, 133.14, 132.21, 131.75, 131.45, 131.16, 130.01, 128.59, 116.27, 55.48, 52.70, 42.22, 37.50; Hi-Res MS calc'd *m/z*, 381.05346 found M-H [380.04781]

(2-(3,4-Dichlorophenyl)acetyl)tyrosine (S4)

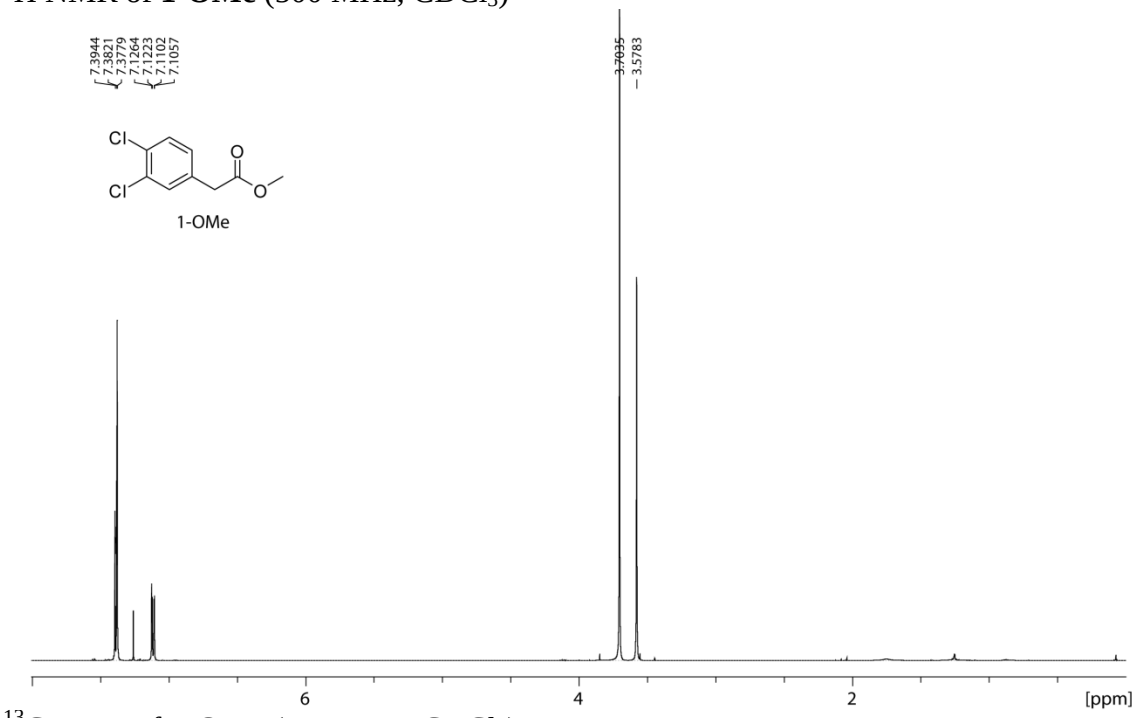
Compound **S4b** was hydrolyzed according to general procedure C on a 0.045 mmol scale. An excess of LiOH was used in this reaction (~400 eq). The final compound **S4** was isolated as a white solid, though a small contamination does not allow for accurate yield calculations. ^{1}H NMR (500 MHz, MeOD): δ 7.40 (d, *J* = 8.3 Hz, 1 H), 7.38 (d, *J* = 2.0 Hz, 1 H), 7.02 (dd, *J* = 8.3, 2.0 Hz, 1 H), 6.97 (d, *J* = 8.5 Hz, 2 H), 6.66 (d, *J* = 8.5 Hz, 2 H), 4.60 (dd, *J* = 9.27, 4.82 Hz, 1 H), 3.47 (d, *J* = 5.61 Hz, 2 H), 3.12 (dd, *J* = 14.0, 4.9 Hz, 1 H), 2.85 (dd, *J* = 14.1, 9.2 Hz, 1 H) ^{13}C NMR (125 MHz, MeOD): δ 175.06 (low intensity), 172.53, 157.33, 137.55, 133.13, 132.26, 131.71, 131.43, 131.23, 130.04, 129.02, 116.19, 55.47, 42.38, 37.60; Hi-Res MS calc'd *m/z*, 367.03781 found M-H [366.03161]

Computation

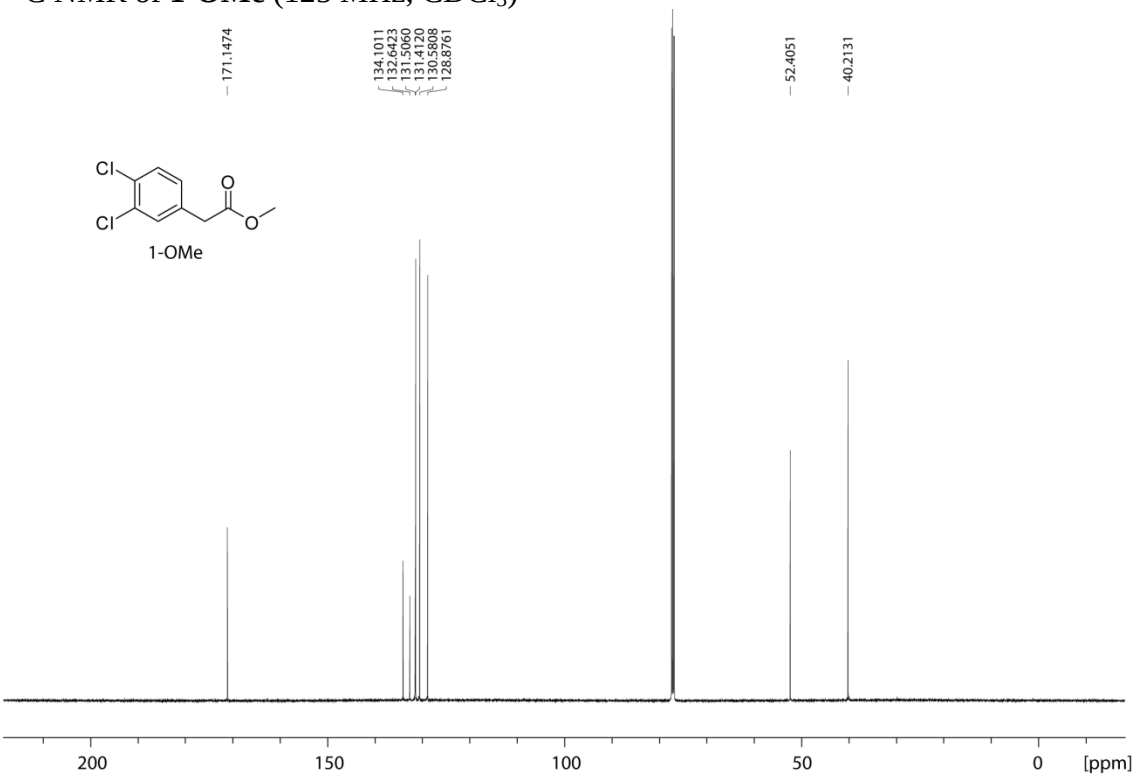
Access to the Schrodinger Software Suite was provided through the Minnesota Supercomputing Institute (MSI). The four following KIX PDB structures were used for docking studies: 1KDX, 2AGH, 2LXT, and 2LXS. The generated grid that was used was 25 Å x 25 Å x 25 Å. Docking studies were performed with extra precision (XP). Two poses for each molecule were generated per PDB structure resulting in eight poses for naphthol AS-E phosphate and eight poses for **2**.

FTMap Computational Solvent Mapping was provided by Boston University at ftmap.bu.edu. [202-205] Experiments were performed on the main KIX chain from each of the four following PDB structures : 1KDX, 2AGH, 2LXT, and 2LXS.

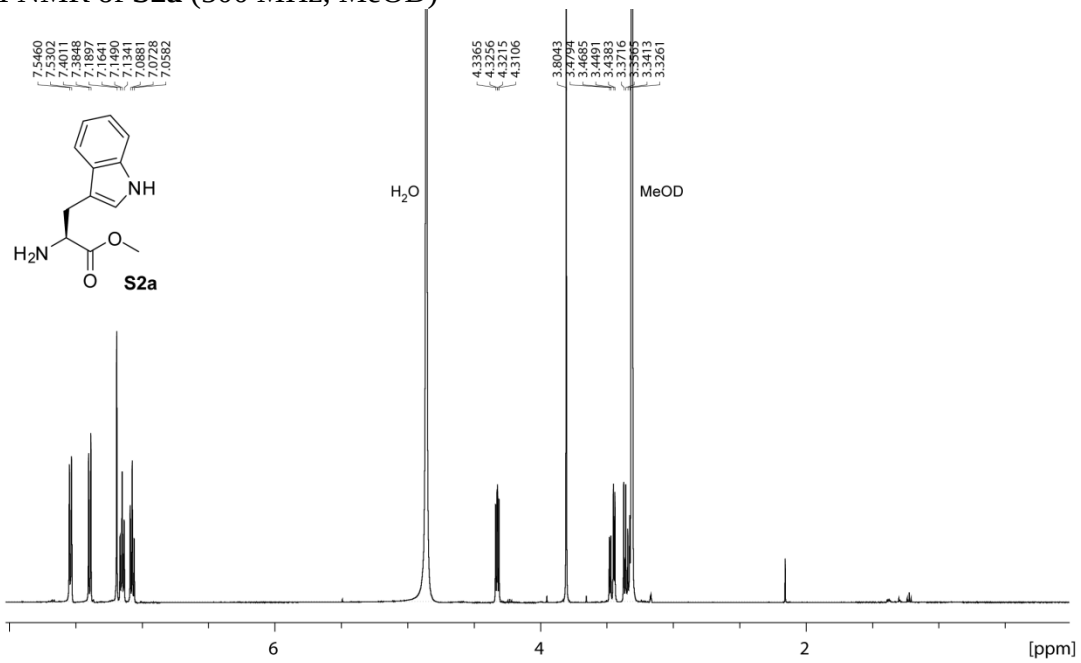
^{1}H NMR of **1-OMe** (500 MHz, $CDCl_3$)



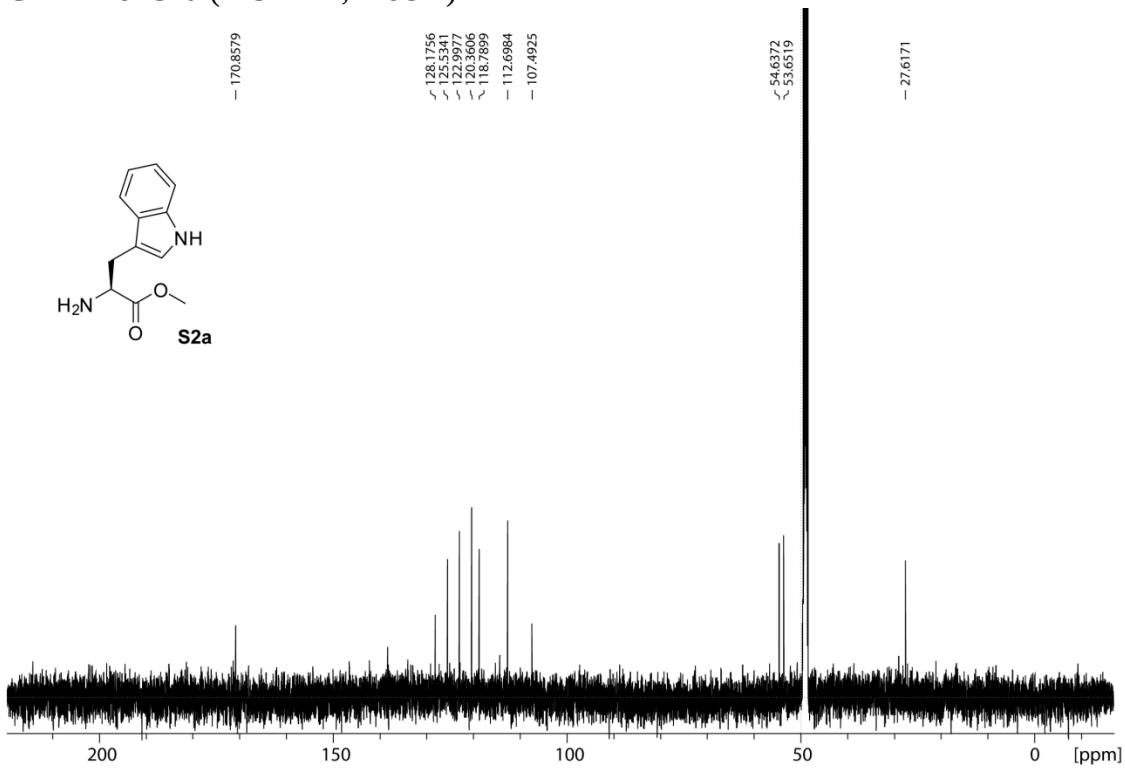
^{13}C NMR of **1-OMe** (125 MHz, $CDCl_3$)



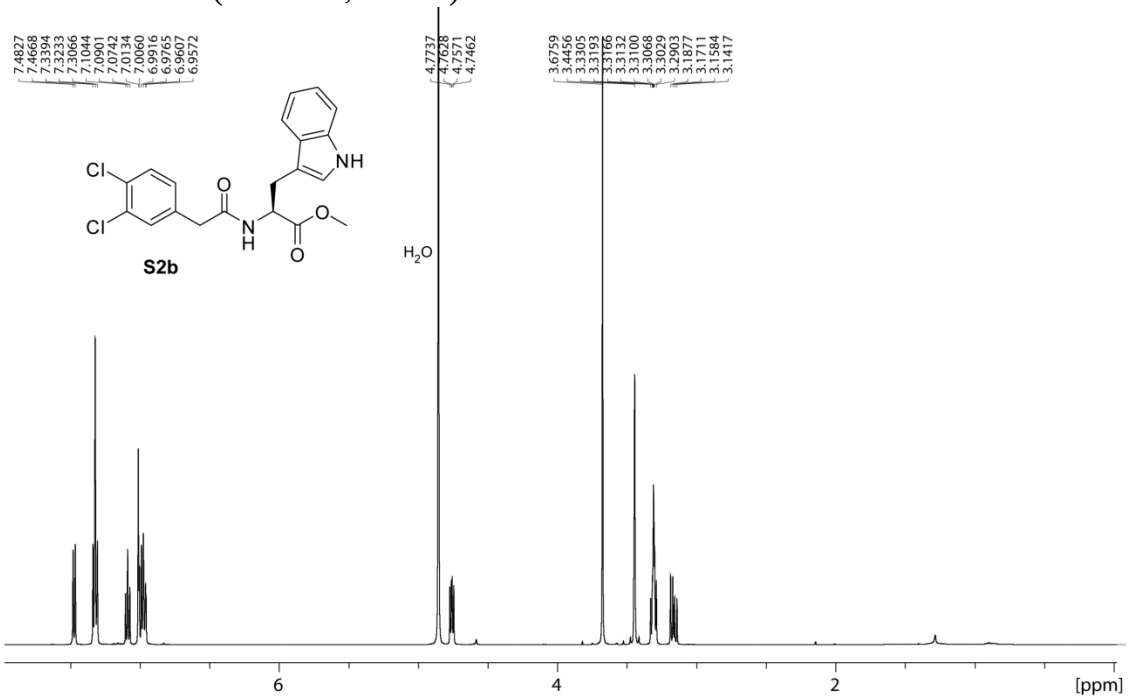
^{1}H NMR of **S2a** (500 MHz, MeOD)



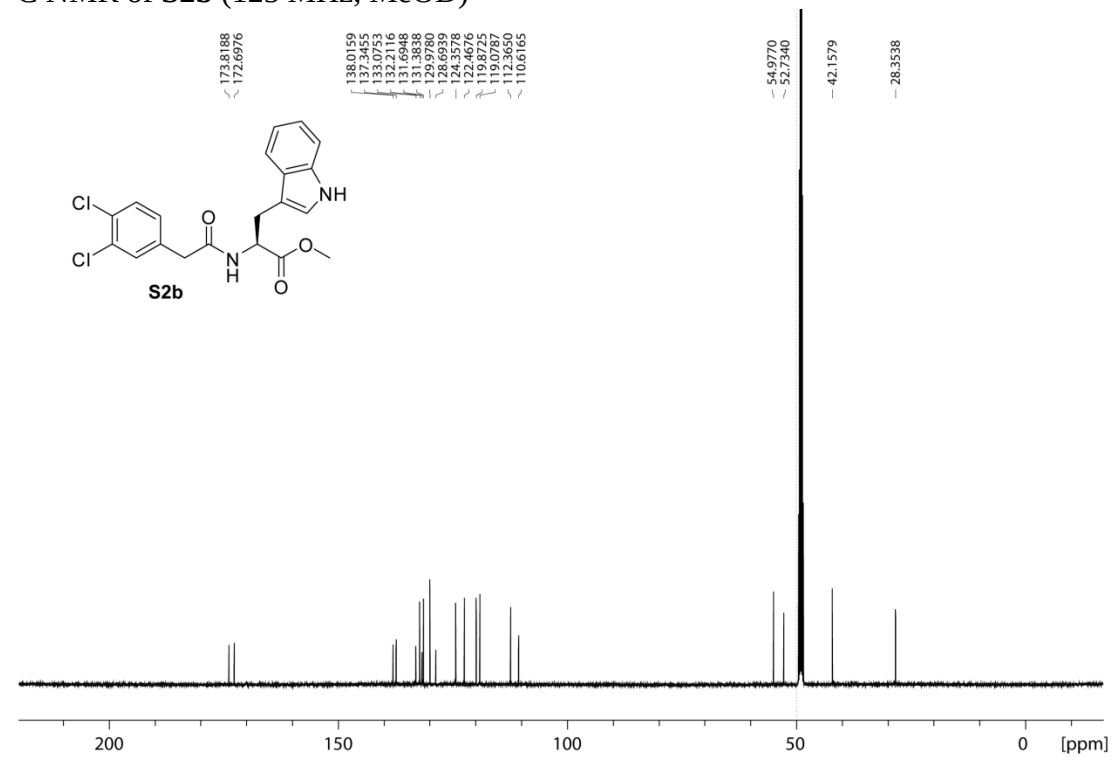
^{13}C NMR of **S2a** (125 MHz, MeOD)



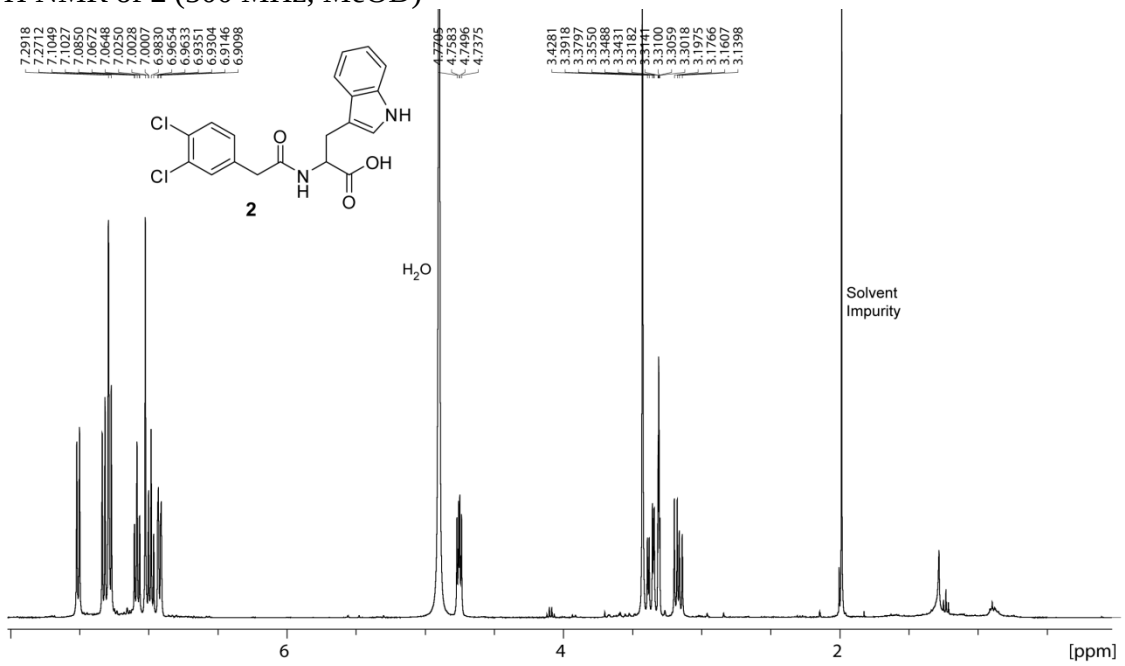
^{1}H NMR of **S2b** (500 MHz, MeOD)



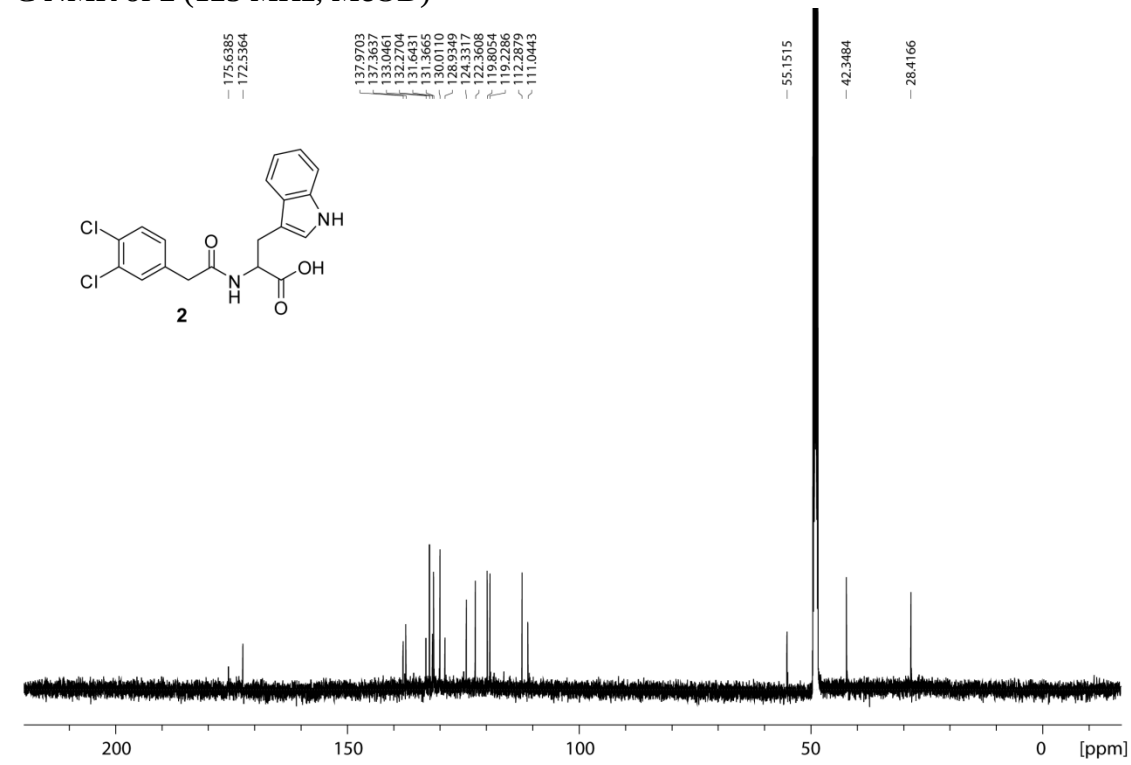
^{13}C NMR of **S2b** (125 MHz, MeOD)



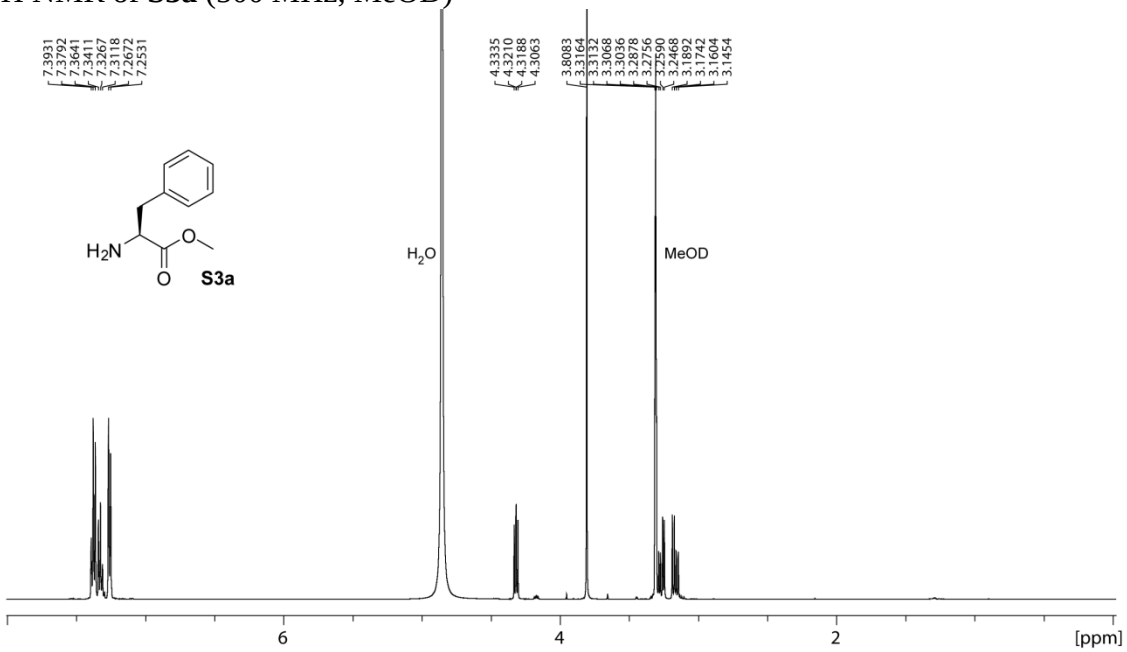
^{1}H NMR of **2** (500 MHz, MeOD)



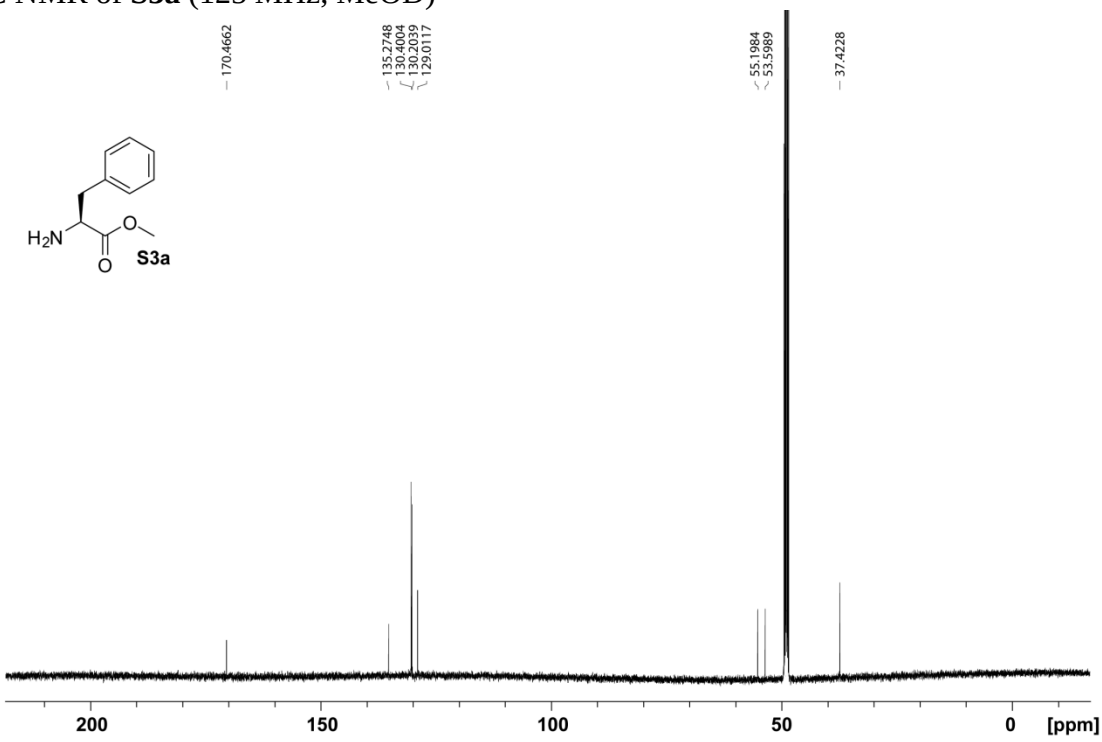
^{13}C NMR of **2** (125 MHz, MeOD)



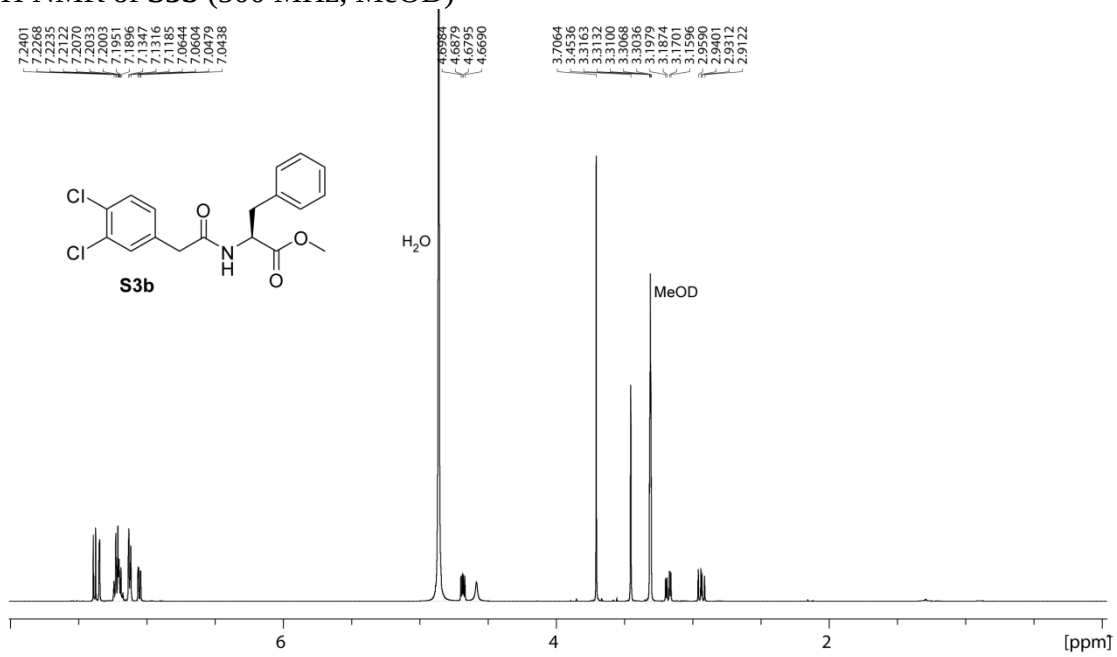
^{1}H NMR of **S3a** (500 MHz, MeOD)



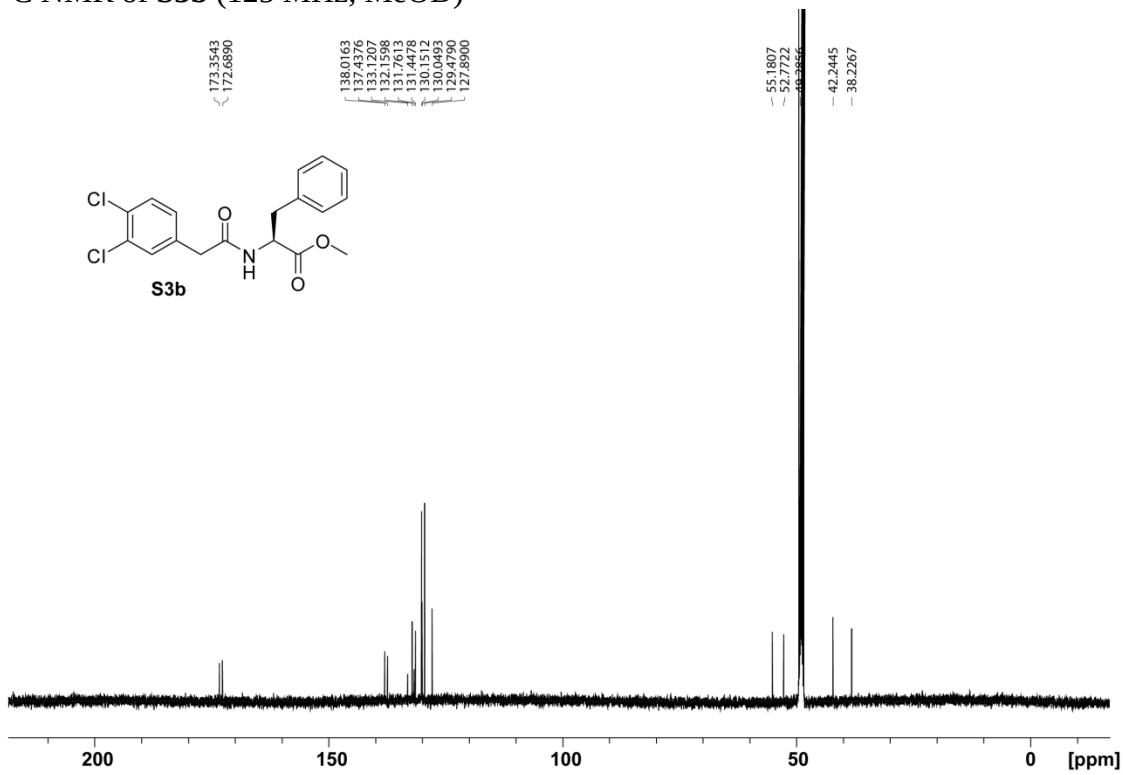
^{13}C NMR of **S3a** (125 MHz, MeOD)



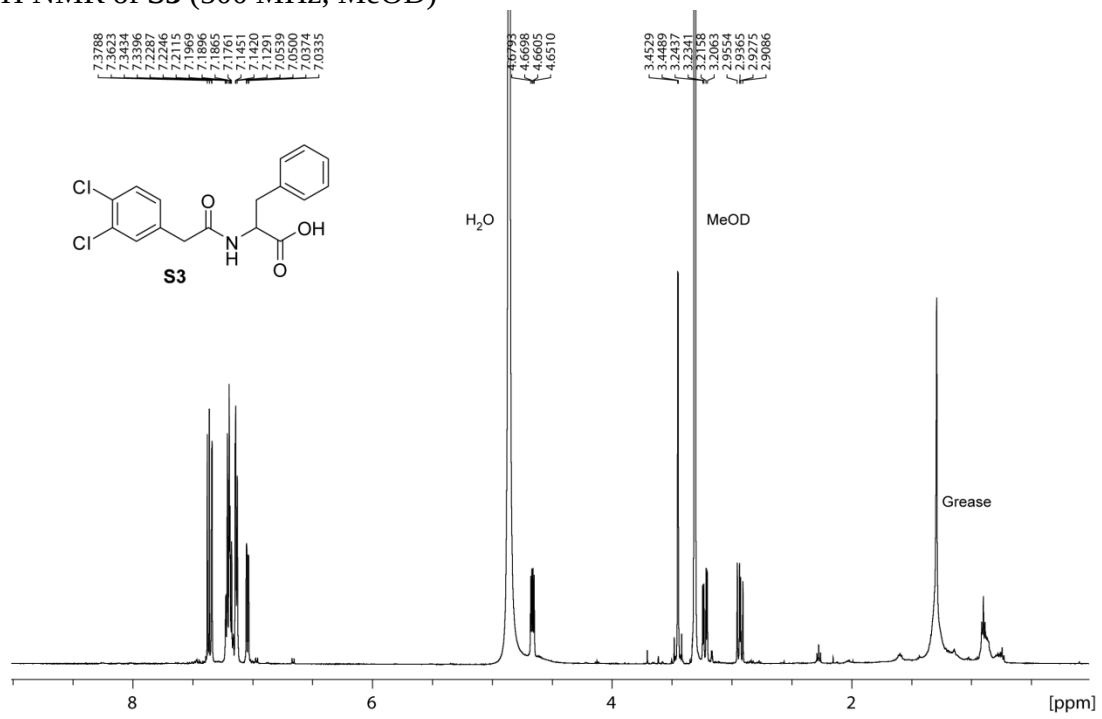
^{1}H NMR of **S3b** (500 MHz, MeOD)



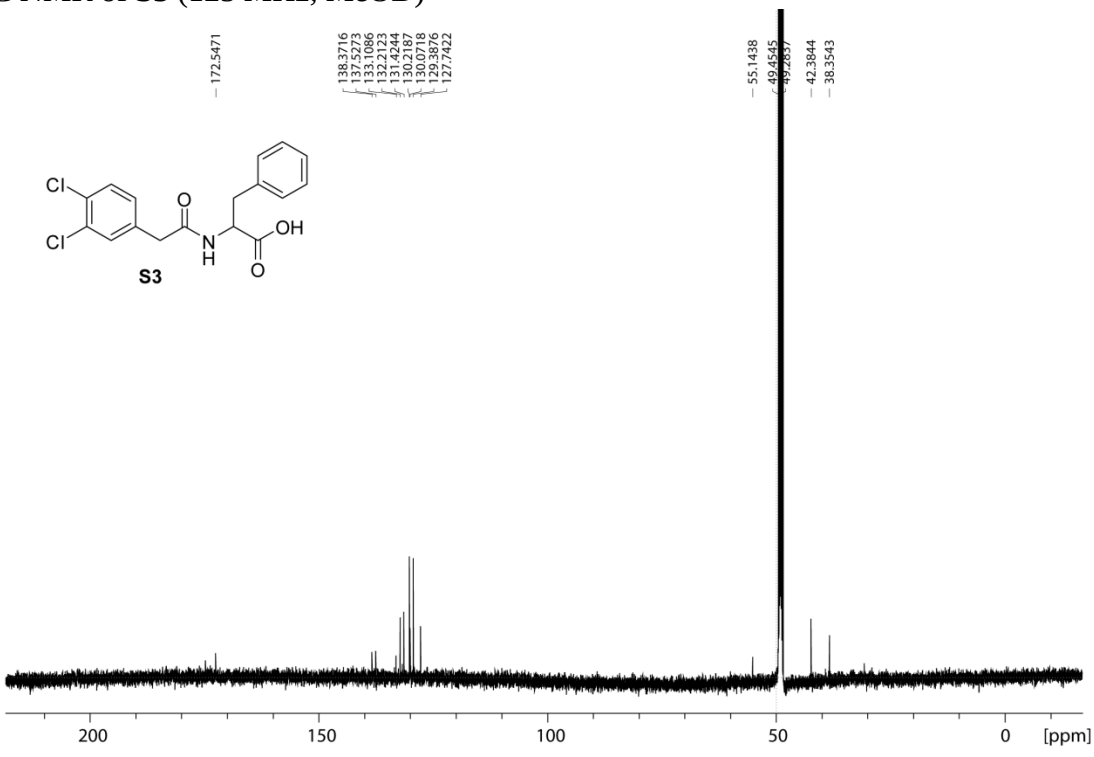
^{13}C NMR of **S3b** (125 MHz, MeOD)



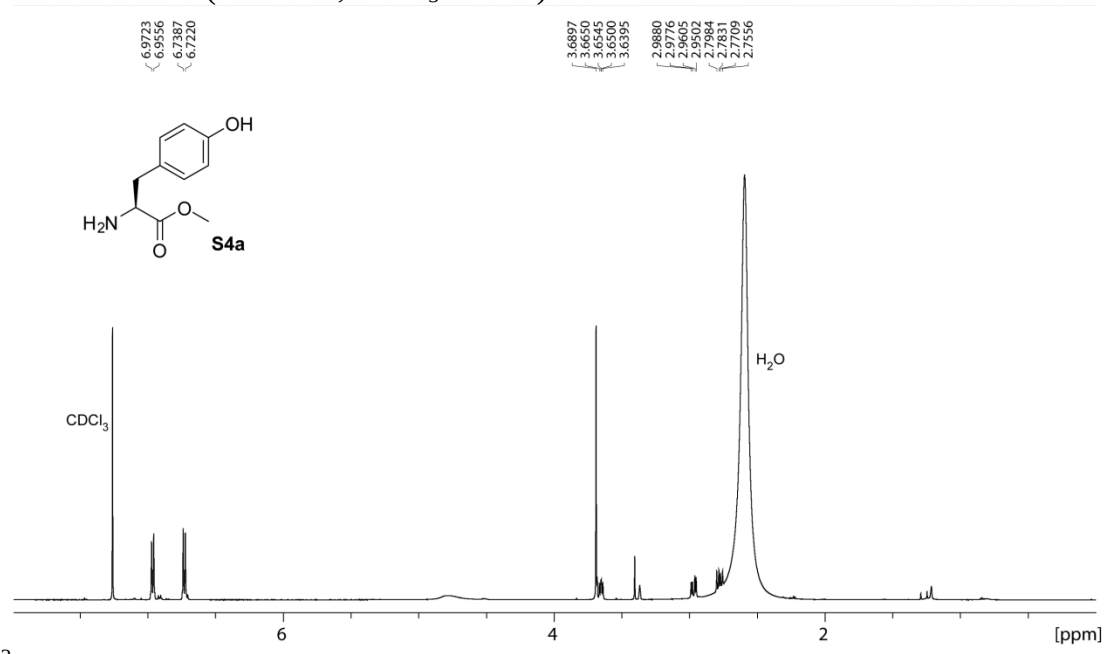
^{1}H NMR of **S3** (500 MHz, MeOD)



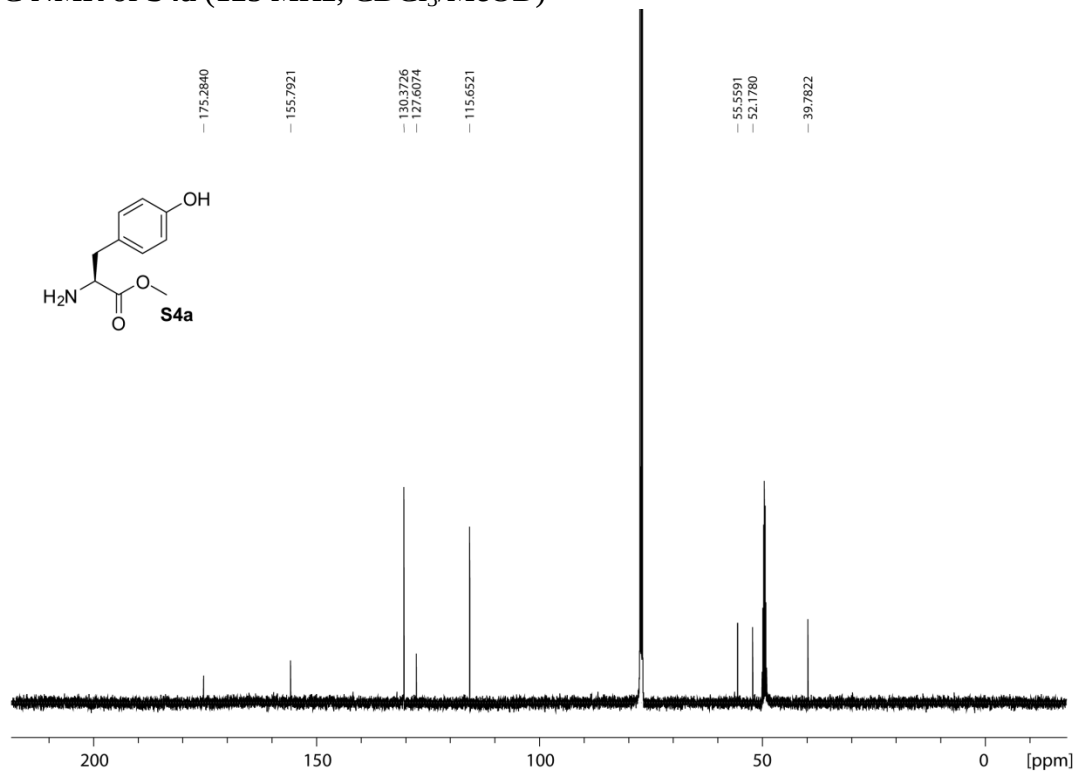
^{13}C NMR of **S3** (125 MHz, MeOD)



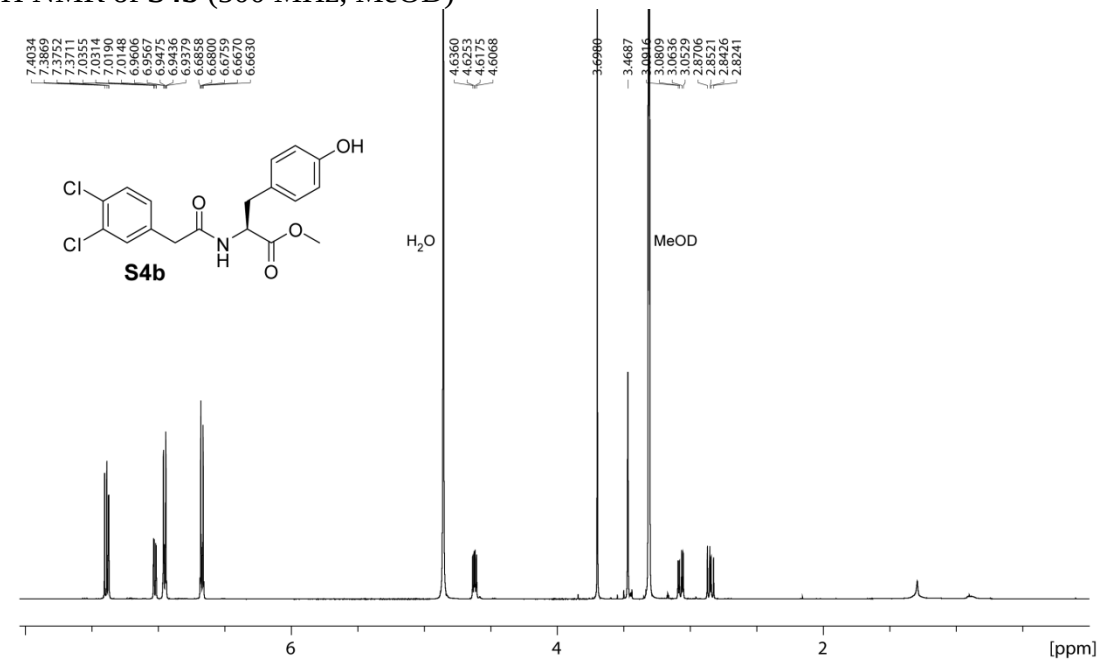
^{1}H NMR of **S4a** (500 MHz, $CDCl_3$/MeOD)



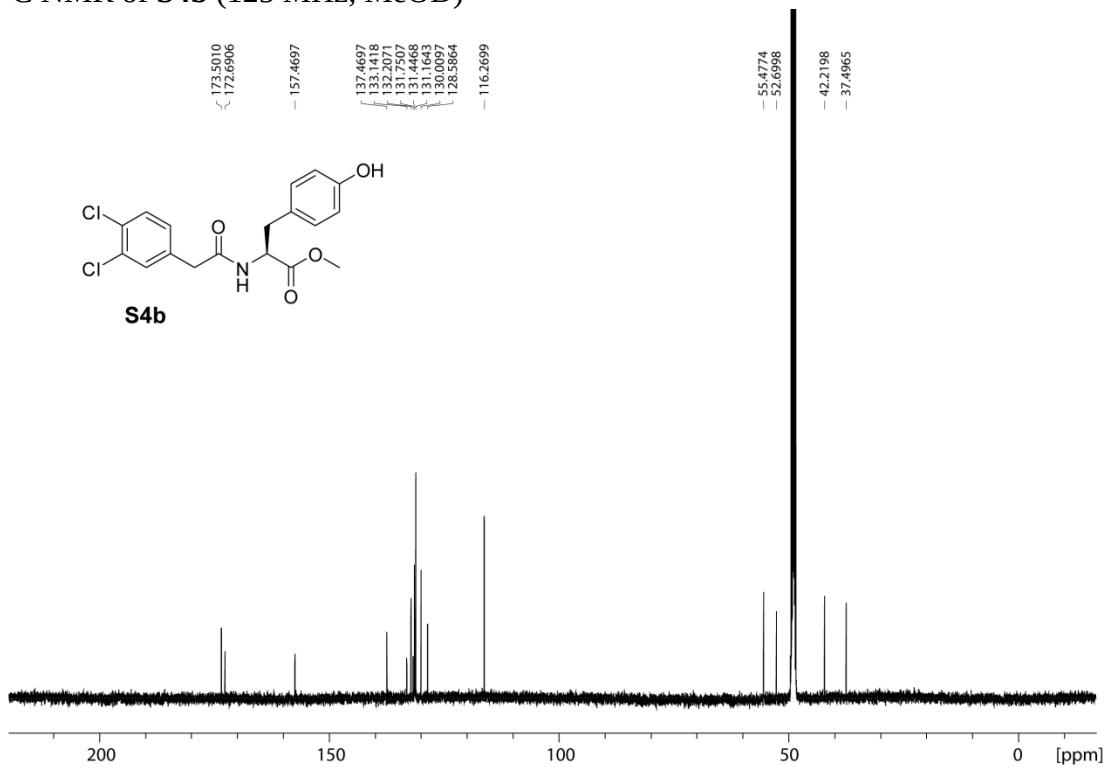
^{13}C NMR of **S4a** (125 MHz, $CDCl_3$/MeOD)



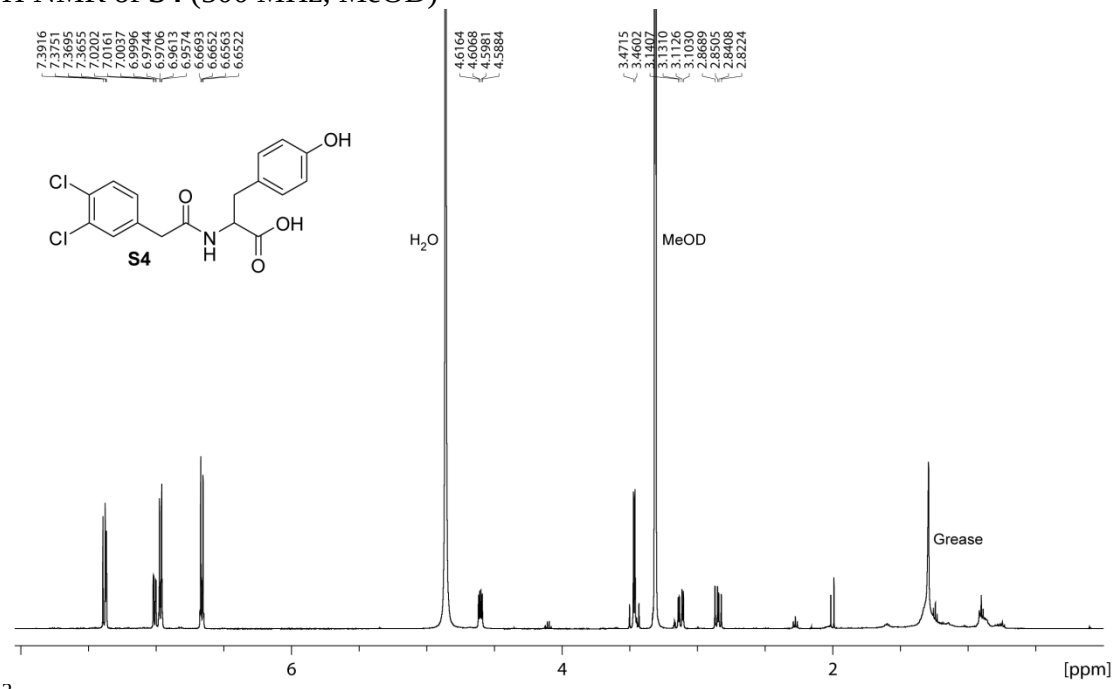
^{1}H NMR of **S4b** (500 MHz, MeOD)



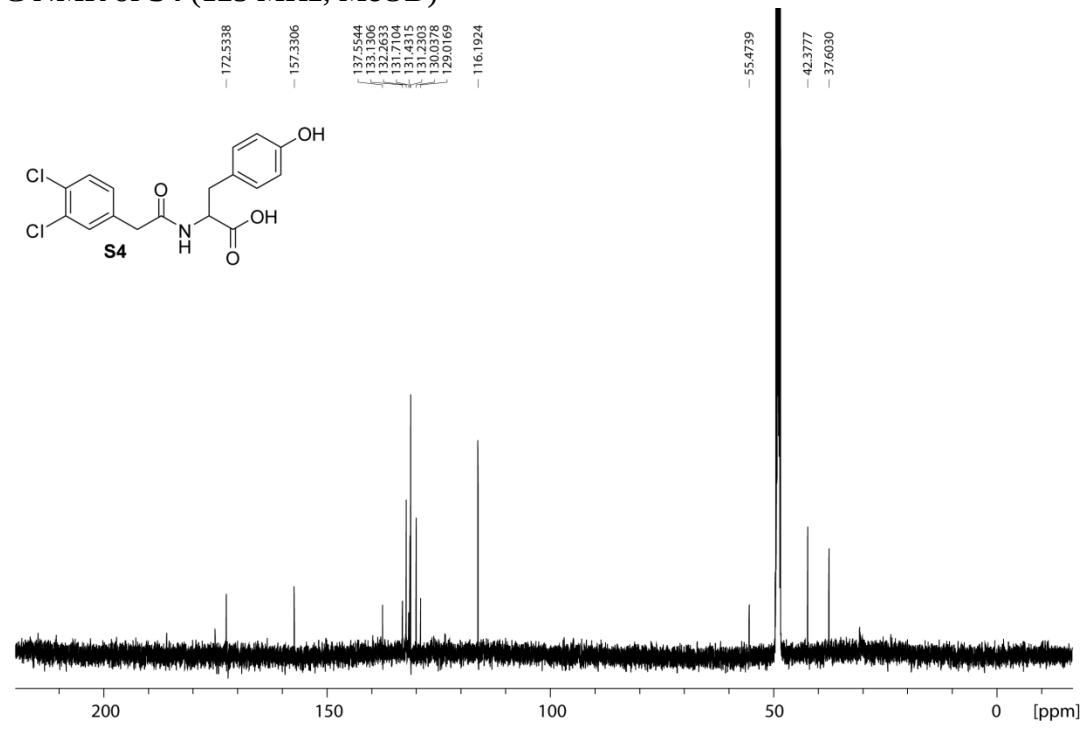
^{13}C NMR of **S4b** (125 MHz, MeOD)



^{1}H NMR of **S4** (500 MHz, MeOD)



^{13}C NMR of **S4** (125 MHz, MeOD)



Chapter 4. Protein-observed 19F-NMR for fragment screening, affinity quantification and druggability assessment

Note: A version of the following was also previously reproduced in the dissertation of Andrew Urick, Ph.D.

Motivation: This manuscript sought to disseminate our procedure for our PrOF NMR method, from protein expression to ligand-binding experiments. The goal of this manuscript was to make it as easy as possible for others to adopt and utilize our techniques for incorporating fluorinated amino acids into proteins of interest for NMR characterization and screening. As such, a paper focused on the details of the methodology was a good fit in order to share both our protocols as well as pitfalls that may be encountered throughout the process.

4.1 **Introduction**

NMR spectroscopy, using either labeled proteins or labeled small molecules is emerging as a preferred method for screening low-complexity molecules (typically < 300 Da with a minimal number of functional groups) termed fragments in early-stage ligand discovery campaigns.[172, 206] Fragments typically bind to their protein target with low affinity (mid micromolar to millimolar dissociation constants). These low affinity interactions are readily detected using NMR methods.[207-208] Fragment molecules can be compared to higher molecular weight counterparts found in traditional high throughput screening libraries, via evaluation of their ligand efficiency, which compares binding affinity or activity relative to the number of atoms in the molecules.[150, 209] Highly ligand

efficient compounds can be developed in an atom-economical manner into more potent compounds by fragment linking or growing approaches.[61, 210] Enthusiasm for this approach remains high with the approval of Vemurefanib[211] in 2011 and Venetoclax[167] in 2016,[212] discovered through an initial fragment screening campaign and several more lead molecules which emerged from fragment screens are in late-stage clinical trials.[147, 211]

Fluorine NMR is an attractive approach for fragment screening because the spin-½ nucleus ^{19}F, is stable, has a natural abundance of 100%, and is nearly absent in biological systems. Many fluorinated amino acids and building blocks are commercially available including aromatic amino acids 3-fluorotyrosine, 4-fluorophenylalanine, and 5-fluoroindole described herein. In many cases, minimal structural and functional perturbation has been observed.[74, 194, 213] ^{19}F chemical shifts are also sensitive to changes in the molecular environment, and therefore ^{19}F is an ideal background-free NMR-active nucleus for studying challenging problems of molecular recognition by biopolymers.[74] [131] In the case of fluorine-labeled proteins, the environmental sensitivity of fluorine nuclei typically results in well-resolved 1D ^{19}F NMR spectra of proteins whose fluorine-labeled side-chains are observed at low-to-mid micromolar concentrations (e.g., 25–100 μM).[74]

Fragment screening using low molecular weight, low complexity molecules has attracted considerable attention due to the reduction of chemical space compared to that of higher molecular weight, functional group-rich small molecules used in high throughput screening. As a result, fragment libraries are typically smaller than high throughput screening libraries[147, 172, 206, 214] An analysis by Scanlon and coworkers[215] of 20 different fragment libraries developed in the context of academic or industrial research yielded an average library member size of 4,543 fragment library members with a median size of 1280**.** The use of smaller library sizes is further supported by the hit rates from these fragment screens where the researchers detected a binding event averaging 8.2% as reported by 11 different screening centers. The high hit rates suggest adequate chemical space is being covered. Fragments identified as hits have been have been used to develop efficient ligands with favorable physicochemical properties.[176]

One of the challenges when screening fragment molecules for binding to a

specific protein target is the detection and quantification of low-affinity interactions. Achieving this important research goal often forces researchers to use high ligand concentrations. NMR is a technique well-suited for working at these high concentrations. With NMR, mixtures of fragment molecules can also be tested simultaneously. In experiments using proteins labeled with NMR-active nuclei, the NMR spectrum of a mixture that results in a large change in chemical shift of the NMR-active nucleus is deconvoluted by obtaining NMR spectra of the protein with individual molecules to find the small molecule that actively binds the protein (thus causing the observed change in chemical shift). One advantage of this fragment mixture approach is that it enables researchers to test a large number of compounds in a shorter period of time than testing individual compounds one at a time. A second advantage when using a labeled protein (i.e., protein-observed NMR methods) is the added structural information the protein resonances provide. These specific resonance perturbations can be used to guide molecular designs aimed at increasing fragment affinity.

Ligand-observed NMR methods, such as saturation transfer difference (STD) or transverse relaxation-Carr-Purcell-Meiboom-Gill-based experiments[216-217], provide complementary information which can be used in parallel with protein-observed methods. Ligand-observed and protein-observed ^{19}F NMR-based methods are tolerant to a variety of experimental conditions. For example, contrary to many ^{1}H NMR-based experiments, NMR spectra of fluorine-labeled small molecules or fluorine-labeled proteins are not affected by the presence of detergents and buffers that are traditionally used.[174] Advantages associated with the implementation of ligand-observed experiments with respect to protein-observed experiments are the lower protein concentration needed (0.1–10 μM) in many cases — although increasing protein concentration can lead to better signal[218] — and lack of an upper limit in protein size. Additionally, with respect to protein-observed experiments, the active ligand can be readily identified from the fragment mixture without deconvolution. However, in STD NMR, false positive hit rates can be as high as 50%,[215] although the occurrence of such false positives can be partly mitigated by repeating the experiment in the presence of a competitor ligand (if one is known).

We have proposed a protein-observed ^{19}F NMR screening method abbreviated as PrOF NMR, in which we rapidly monitor changes in the chemical shifts of fluorine resonances in ^{19}F-labeled protein side chains induced by the presence of small molecules Figure 4.1). An approach that is analogous to that adopted in structure–activity relationship (SAR) by NMR using labeled amides in ^{1}H-^{15}N heteronuclear single quantum coherence spectroscopy (HSQC) experiments.[61] The increasing availability of improved instrumentation using ^{19}F-tuned cryoprobes, (e.g., the QCI-F cryoprobe from Bruker), speed of PrOF NMR experiments, and ease of spectral interpretation, have increased the accessibility of these experiments in academic and industrial settings. We recently applied PrOF NMR for fragment screening with the transcription factor–binding domain KIX.[41, 193] In this study, we analyzed in 85 mixtures (comprising a total of 508 small molecules) in 10 hours of experiment time using a total of 20 mg of protein. PrOF NMR has been employed by others, including follow-up screens against the SPRY domain-containing SOCS box protein 2[219] and AMA1[220], and shown to be more than twice as fast as ^{1}H-^{15}N HSQC NMR for small proteins.[193, 221] We have also used the bromodomain BrdT, described below.[98] G-protein coupled receptor (GPCR) agonists and antagonists can also be identified by this method.[79] 2D PrOF NMR methods have validated binding modes of fluorinated ligands via ^{19}F–^{19}F homonuclear nuclear Overhauser effect (NOE) experiments with BcL-xL.[222] As the incorporation of fluorine in drug molecules increases, NOE experiments provide an additional structural biology tool for characterizing ligand-binding modes. Our lab recently demonstrated a simultaneous analysis (i.e., multiplexed) of small-molecule binding using two 15-kDa bromodomains Brd4 and BPTF[98] In this study, 229 small molecules were screened using this approach.[99] Because a protein and potential off-target are screened together, this multiplexed experiment is similar to the RAMPED-UP 2D NMR experiments with differently labeled proteins.[223] These approaches are advantageous for studies where protein selectivity is important.

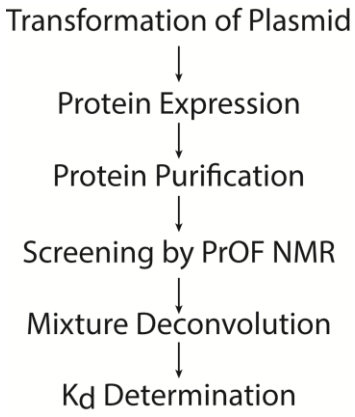


Figure 4.1 General workflow for using PrOF NMR for fragment screening and ligand characterization.

In 1974, Sykes et al. first reported the ^{19}F NMR analysis of an 86-kDa protein, alkaline phosphatase, using 3-fluorotyrosine (3FY) to label the protein.[137] Sequence-selective labeling of recombinant proteins with fluorinated amino acids is now a well-documented methodology that facilitates the general use of the present screening method based on PrOF NMR.[200, 224] In this protocol, using sequence-selective labeling with fluorinated aromatic amino acids, we describe the application of PrOF NMR to the screening of small-molecule libraries as potential protein ligands and the quantification of the micromolar to millimolar dissociation constants from chemical shift perturbation analysis. We demonstrate our ligand-binding screening method on two proteins, the transcription factor–binding domain of the CREB-binding protein, KIX, and the bromodomain Brd4 (Figure 4.2). Application of PrOF NMR with a third protein, the first bromdomain of BrdT, will be subsequently described to highlight several additional important aspects of the PrOF NMR protocol. The sequence-selective incorporation of the three fluorinated aromatic amino acids, 3FY, 4-fluorophenylalanine (4FF), and 5-fluorotryptophan (5FW) into recombinant proteins in *E. coli* using either the auxotrophic bacterial cell lines [e.g., DL39(DE3)] or standard bacterial strains [e.g., BL21(DE3)] will first be detailed, as previously described.[41, 200, 224] This section of the Procedure will then be followed by our ligand-discovery procedures.

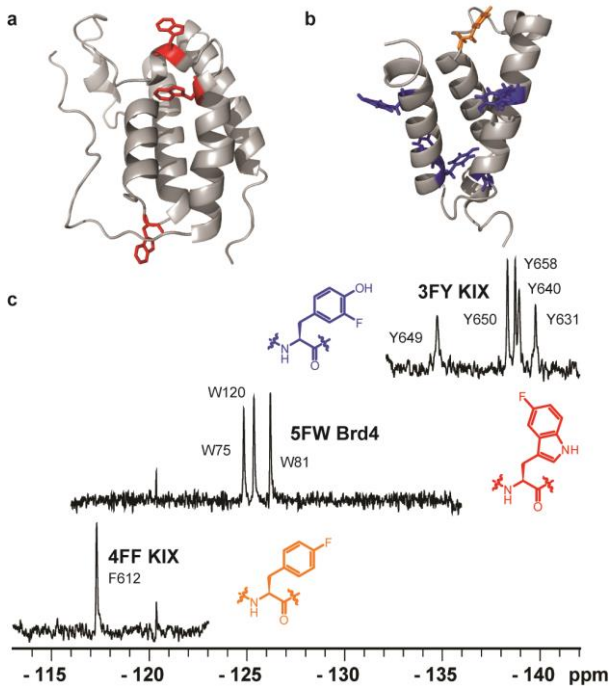


Figure 4.2 PrOF NMR examples.
(a) Crystal structure of Brd4 (PDB 3MXF) with its three tryptophan residues shown in red. (b) Solution structure of KIX (PDB 1KDX). (c) PrOF NMR spectra for each of these variants, in which the residues highlighted in **a** and **b** were replaced by 5FW (Brd4), 3FY (KIX) or 4FF (KIX). In these spectra, the wide chemical shift dispersion of the signals due to the fluorine-labeled, aromatic amino acid analogs are highlighted. The structures of the mentioned fluorinated amino acids are shown next to their corresponding spectra.

4.2 **Experimental Design**

4.2.1 Fluorinated Protein Expression and Characterization.

All expressed proteins in the Procedure are characterized for purity and fluorine incorporation via SDS-PAGE and protein mass spectrometry via electrospray ionization. Protein yields vary based on the protein system, fluorinated amino acid, and cell line used. We have achieved yields as high as 70 mg/l for 3FY-labeled KIX and 65 mg/L for 4FF-labeled KIX using auxotrophic DL39(DE3) cells, while obtaining up to 62 mg/L of 5FW-labeled KIX[23] using non-auxotrophic BL21(DE3) cells with 5-fluoroindole added to the cell culture medium. All expressions led to high labeling efficiency. We observed even higher yields with our fluorinated bromodomains (84 mg/L for 5FW BPTF and 88 mg/L for 5FW Brd4).[99] The SDS-PAGE gel in Figure 3A shows four KIX protein

samples that are either unlabeled, 3FY-labeled, 4FF-labeled, or 5FW-labeled.

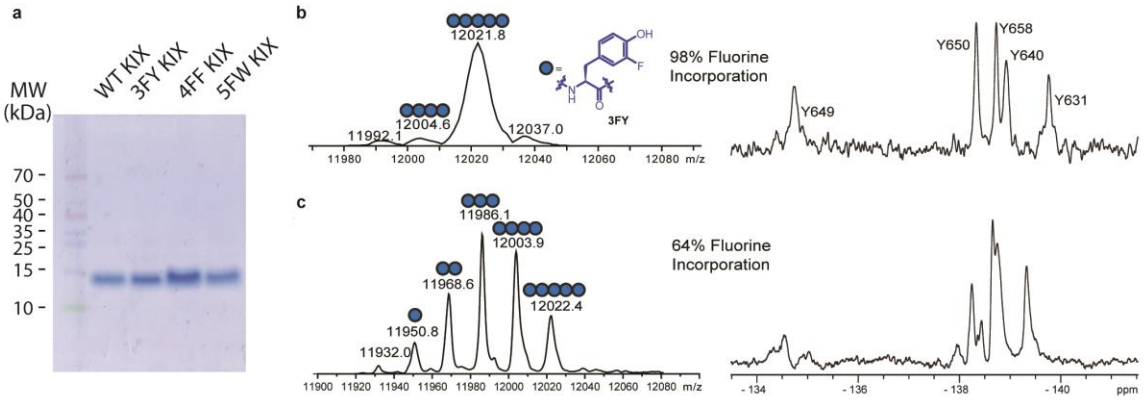


Figure 4.3 Characterization of fluorinated proteins expressed according to present protocol.
(a) SDS-PAGE gel stained with Coomassie blue showing wild-type and three fluorinated variants of KIX. (b) Deconvoluted electrospray ionization mass spectrometry (ESI-MS) spectrum (left) and ^{19}F NMR spectrum (right) of 3FY-labeled KIX with 98% fluorine incorporation. The dominant population present is the fully fluorinated variant. (c) Deconvoluted ESI-MS spectrum (left) and ^{19}F NMR spectrum (right) of 3FY-labeled KIX with 64% fluorine incorporation.[193]

Mass spectrometry is used to assess fluorine incorporation into the protein. For the mass spectrogram of 3FY-labeled KIX, the dominant mass of 12,021.8 Da corresponds to the incorporation of a fully labeled protein (five 3FY residues). The minor mass of 12,004.6 Da corresponds to four of the five tyrosine residues being replaced with 3FY. These major and minor populations lead to a 98% labeled protein Figure 4.3 b). Low levels of incorporation can result in a heterogeneous protein sample, which can complicate the analysis of ^{19}F NMR spectra (see troubleshooting sections). A ^{19}F NMR spectrum of 3FY-labeled protein that is only 64% labeled is reported for comparison (Figure 4.3 c). The resulting spectrum is a statistical mixture of multiply labeled proteins and can result in additional (e.g. Y650) or broadened resonances. In some instances, and if needed, the concentration of fluorinated amino acids can be increased in the culture media with respect to those recommended in the Procedure, to increase the extent of protein labeling. Low levels of labeling can also be attributed to residual unlabeled amino acid from the initial expression conditions, an inconvenience that can be reduced with careful washings of bacterial pellets and preincubations with fluorinated amino acids.

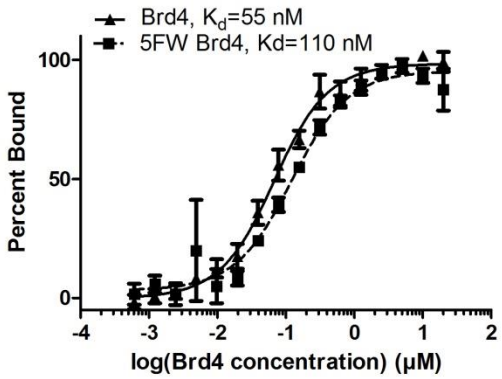


Figure 4.4 Direct binding experiment of BODIPY-BI, a Brd4 bromodomain ligand to unlabeled Brd4 and 5FW-labeled Brd4 by fluorescence polarization.

Fluorine is not an element naturally found in proteins. Therefore, prior to the implementation of the screening protocol, the structural and functional perturbations caused by the introduction of the fluorine label in the protein must be carried out. We have characterized our proteins by a variety of methods, using x-ray crystallography, circular dichroism and thermal stability measurements to assess structure.[41, 98] When a ligand was known beforehand, we have also employed isothermal titration calorimetry or fluorescence anisotropy ligand-binding experiments to compare the affinity of the known ligand with the fluorinated and non-fluorinated protein. Figure 4.4 shows an example of a direct binding experiment with a fluorescently-labeled bromodomain ligand, BI-BODIPY with 5FW-labeled Brd4 and unlabeled Brd4 yielding dissociation constants of 110 nM and 55 nM, respectively. In our experience, we have considered a two-to-three-fold change in binding affinity between fluorine-labeled and unlabeled protein acceptable for continuing on with the labeled protein with a ligand screen. We recommend alternative labeling approaches or amino acids (e.g., 6-fluoro versus 5-fluorotryptophan) to be tested if larger perturbations are observed.

4.2.2 Protein-observed fluorine NMR.

PrOF NMR spectra of small to medium-sized fluorinated proteins typically reveal well-resolved resonances for each labeled aromatic amino acid, but not in all cases. The different aromatic residues are seen in close but distinct chemical shift regions (Figure 4.2). A well-dispersed NMR spectrum is an additional confirmation of a well-folded

protein. Protein resonances tend to be broad. Fluorine resonances from small-molecule impurities or peptides resulting from proteolytic degradation appear as sharp signals in ^{19}F NMR spectra, and their presence can help assess the stability and purity of the protein. A resonance due to a small-molecule impurity can be seen in Figure 4.2 at –120 ppm. In some cases, protein precipitation or aggregation leads to loss of signal for all of the protein resonances, which can be useful for identifying false positives in a ligand screen.

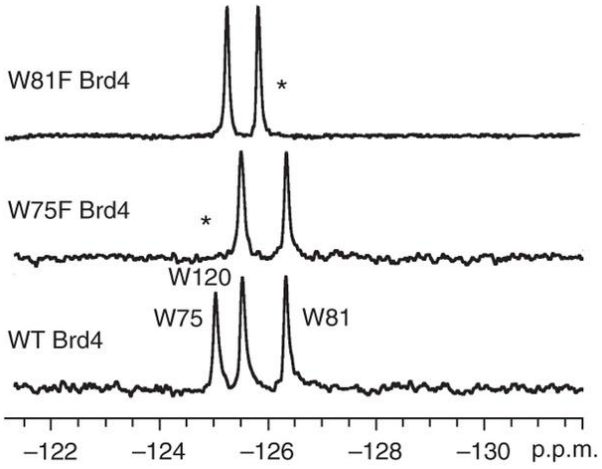


Figure 4.5 Using site-directed mutagenesis to assign PrOF NMR resonances. Asterisks denote the resonance that has disappeared after mutation of the relevant (fluorine-labeled) tryptophan into a phenylalanine, which allows assignment of that resonance to the specific mutated amino acid residue.[98]

Resonances can be assigned by recording the spectrum of a fluorinated protein, where one residue has been mutated to an amino acid that can no longer be fluorine-labeled in the expression conditions, because the corresponding fluorinated analog is not present in the culture medium (e.g., Trp to Phe). As an example, three relevant spectra are reported in Figure 4.5. In this case, the bromodomain Brd4 is labeled with three 5FW residues, which results in three well-resolved resonances. The disappearance of the resonances at –124.9 ppm and –126.2 ppm upon directed mutation of specific tryptophan residues leads to the assignment of these resonances to fluorine-labeled W75 and fluorine-labeled W81, (Figure 4.5, middle and top spectra). However, due to the sensitivity of fluorine to subtle changes in environment, additional chemical shift changes can occur, which, in some cases, preclude assignment. In this case, addition of a known

ligand can help identify side-chain resonances in the ligand-binding site. For example, the first bromodomain of BrdT has two tryptophan resonances, and W50 is known to be located in the WPF shelf at the histone-binding site. Addition of the known BrdT ligand dinaciclib (IC_{50} = 61 μM) was used to help assign the resonance due to the fluorine-labeled W50. The broadening and the 0.25-ppm chemical shift perturbation of the upfield resonance is consistent with x-ray crystallography data on the co-crystal structures, which show a significant conformational change of W50 upon binding[225] (Figure 4.6 and Figure 4.7). This assignment was also confirmed by site-directed mutagenesis.[166] Chemical shift perturbation upon addition of a known ligand can also serve as a useful confirmation of the functional integrity of a fluorine-labeled protein.[41, 98] Alternative strategies for assignments are described in the troubleshooting section. Fragment screening can be carried out once a well-folded highly labeled protein has been obtained, although the protein NMR spectrum does not necessarily need to be assigned first.

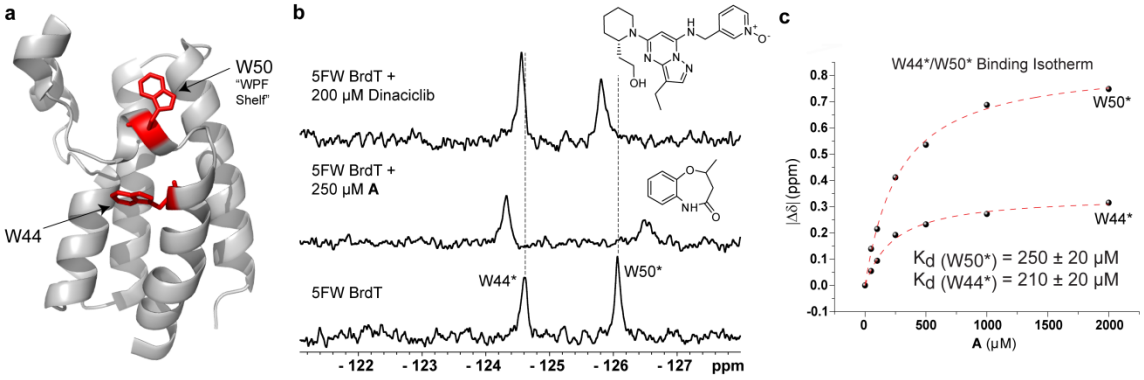


Figure 4.6 Results from experiments involving small-molecule ligands for BrdT.
(a) Crystal structure of BrdT (PDB 4FLP) showing the two tryptophans. Please note that W50 is located in the protein's WPF shelf. (b) PrOF NMR spectra of 5FW-BrdT by itself (bottom spectrum) and in the presence of two different small molecules, fragment A, middle spectrum (see also the structure of the compound above the spectrum, to the right-hand side), and dinaciclib, top spectrum (see also the structure of the compound above the spectrum, to the right-hand side). (c) Binding isotherms for fragment A generated by plotting the change in chemical shift for both 5FW resonances as a function of ligand concentration, which yield comparable K_d values.

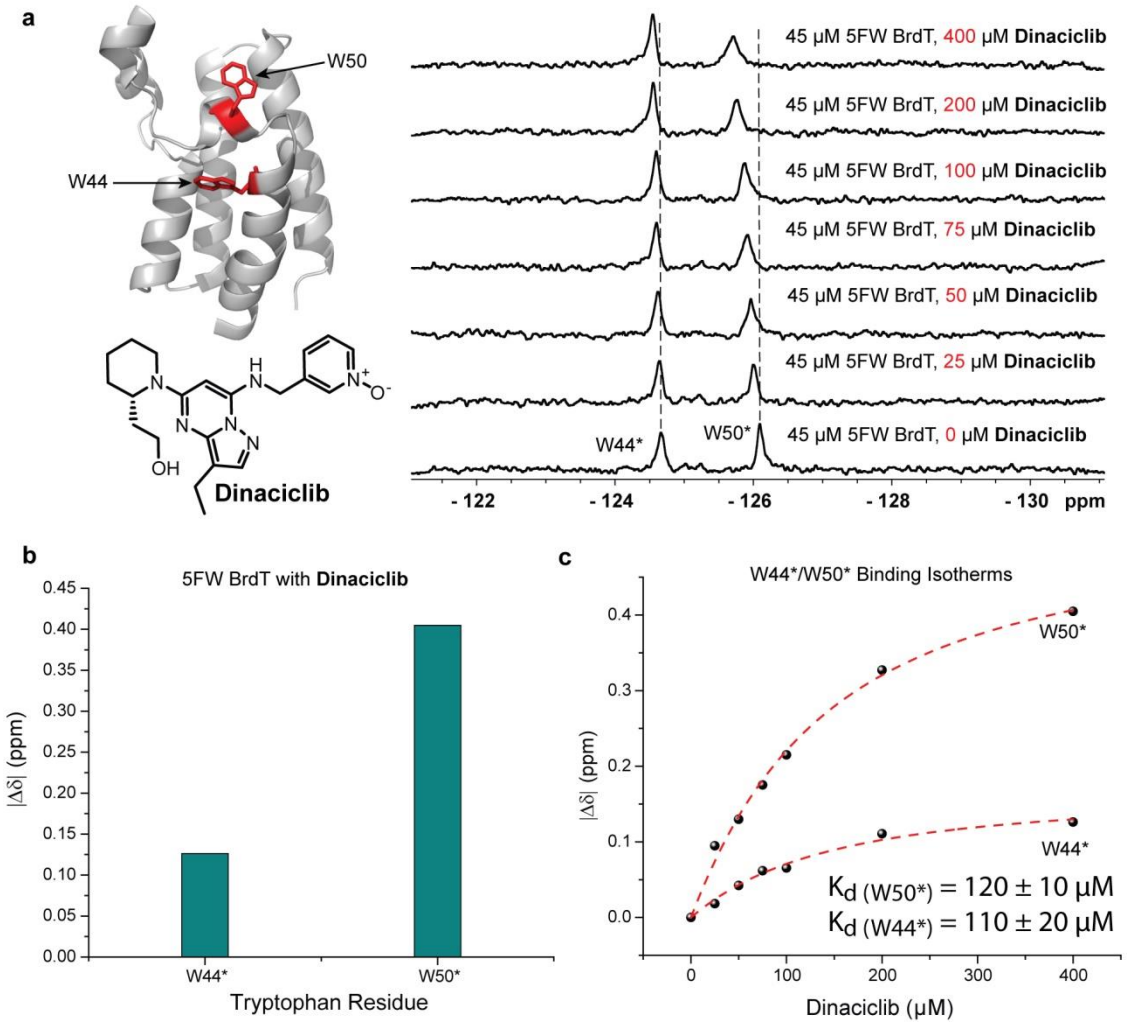


Figure 4.7 Titration of Dinaciclib with BrdT.
(a) Left: Ribbon diagram of BrdT (PDB Code: 4FLP) with tryptophan side chains indicated as sticks shown in red. Right: Stacked PrOF NMR spectra with increasing concentration of dinaciclib. (b) Absolute value of chemical shift perturbations for all 5FW BrdT tyrosine resonances at 400 μM Dinaciclib. (c) Binding isotherm of both W44 and W50 perturbations for the titration with Dinaciclib.

4.2.3 Dissociation constant and ligand efficiency determination.

The affinity of fragment hits from the screen can be readily assessed for protein ligands with up to millimolar dissociation constants via ligand titration and subsequent non-linear regression of the binding isotherm produced by monitoring changes in chemical shift. This chemical shift perturbation experiment is valid when the molecules

are in fast exchange, as indicated by the presence of a single resonance signal, whose chemical shift is the weighted average of the bound and unbound states. The speed of the experiment, the low concentration of proteins needed for it, and the possible automation of the procedure enable the rapid testing of multiple small molecules.

In the present approach, separate samples are set up with varying concentrations of the small molecule. The dissociation constant (K_d) is obtained by fitting the obtained data to a non-linear regression curve using equation 1, accounting for receptor depletion.[221] Analysis of the affinity of the small molecule and location of perturbed resonances can then be used to assess valuable structure–activity relationships between molecules and differential binding to alternate surfaces on a protein.

$$S = D * \frac{(K_d+L+P)-\sqrt{(K_d+L+P)^2-4PL}}{2P} \quad (1)$$

In equation 1, S is the observed change in chemical shift, D is the maximum change in chemical shift, K_d is the dissociation constant of the ligand, L is ligand concentration, and P is protein concentration. The maximal shift, D, can be obtained through non-linear regression, or experimentally, when further addition of a small molecule no longer perturbs the resonance. Additionally, as a means of prioritizing ligands for future development, these affinity data can be used to calculate the ligand efficiency (LE) using eq. 2, which enables this efficiency to be calculated on the basis of the binding affinity of the ligand and the contribution of each non-hydrogen atom to the free energy of the interaction.

$$LE = \frac{-RT\ln(K_d)}{\#\ Non{-}hydrogen\ atoms} \quad (2)$$

Although this scenario is not as commonly encountered in fragment screens, ligands with high affinity may have a sufficiently long residence time on the protein that the bound and unbound states are partially resolved (intermediate exchange) or fully resolved (slow exchange) in the ^{19}F NMR experiment. In these instances, titration will not result in a binding isotherm. Intermediated exchange kinetics result when interchanging species at equal populations (e.g., bound and unbound states) lead to coalescence into a single resonance.[226-227]

$$\tau_{coalescense} = (\sqrt{2}\pi\Delta\upsilon)^{-1} = k_{-1} \quad (3)$$

In equation 3, the frequency difference in Hertz between the two resonances (Δν) — bound versus unbound state — establishes a relationship that enables researchers to estimate the residence time of the ligand on the protein based on the rate of dissociation (k_{-1}). In the case in which the rate of exchange approaches the coalescence point ($\tau \approx \Delta\nu$), during a protein–ligand titration, the resonances will coalesce, so that the signal will initially broaden, but it will become sharp again, in a dose-dependent manner, as the chemical shift is perturbed by the titration process. With slow exchange kinetics (likely to be associated with a low dissociation constant), both resonances can be well resolved, and, during a protein–ligand titration, one resonance will gradually disappear while another will grow from the baseline in a dose-dependent manner.

4.2.4 Experimental considerations and assay limitations

Although the additional rotational freedom in protein-side chains versus the amide backbone can lead to smaller linewidths relative to amide resonances. One consideration that should be made before using ^{19}F-labeled proteins for NMR is the sensitivity of the fluorine nucleus to substantial chemical shift anisotropy (CSA) relaxation effects. CSA relaxation leads to resonance broadening, which is enhanced by protein dynamics, including long rotational correlation times of amino acids found in large proteins.[125] CSA relaxation is proportional to the square of the magnetic field strength. From a practical standpoint, we carry out NMR experiments at 471–564 MHz using fluorinated aromatic amino acids. 5FW experiences a lower degree of CSA relative to other fluorinated tryptophan analogs, and in some cases it has been used with proteins as large as 65 kDa.[228] As an alternative to 5FW, fluorine-labeled cysteine or methionine derivatives with substantially reduced CSA effects are recommended for proteins larger than 40 kDa, as has been demonstrated in experiments with GPCRs and amyloid-β aggregates.[78-80, 229] Site-selective fluorine labeling with trifluoromethylphenylalanine has also been employed for large proteins.[230-231] For small-to-medium-sized proteins, PrOF NMR is applicable using a variety of fluorine-labeled amino acids, including singly fluorinated (tyrosine, phenylalanine, and tryptophan, Fig. 2) as well as more heavily fluorinated amino acids, such as di- and trifluoromethylmethionine[140, 232] and tri- and

hexafluoroleucine[233-235]. We propose using singly fluorinated aromatic amino acids, as this approach reduces spectral complexity with respect to using amino acids with multiple non-equivalent fluorines. Furthermore, it allows high levels of enrichment in labeled amino acids at protein–protein interaction interfaces,[170] while minimizing the perturbing effect of fluorine substitution. Nevertheless, we reiterate that protein structural and functional perturbation caused by fluorine substitution should always be assessed conducting preliminary complementary structural and functional biophysical experiments.

4.3 **MATERIALS**

4.3.1 Reagents:

Unless noted, reagents of comparable quality from alternative suppliers can be substituted.

Protein Expression Components

- Fragment mixtures. <CRITICAL> These are prepared according to guidelines outlined in Box 1. Sources of fragment libraries include Maybridge and Chembridge
- Plasmid encoding your protein of choice that also encodes for the resistance to an antibiotic of choice

Critical: For nickel affinity purification**,** ensure that the plasmid includes a His tag described in step 9 and in the supporting information.

- Plasmid vectors pRARE (Novagen), pRSETB-HIS6KIX (Invitrogen), pNIC28-BSA4 (Addgene)
- Competent DL39(DE3)* (DL39, CGSC) or BL21(DE3) (Novagen) *Escherichia coli* cell lines
  Note: We thank I. Ropson for providing the DL39(DE3) cell line.

Critical: Choice of cell lines must be made carefully. Auxotrophic cell lines (e.g., DL39(DE3)) are necessary for labeling with 3-fluorotyrosine or 4-fluorophenylalanine. For tryptophan labeling, 5-fluoroindole with non-auxotrophic cell lines (e.g., BL21(DE3)) is recommended for higher protein yields.

- QiAprep Spin Miniprep Kit (250) (Qiagen, Cat. No. 27106)
- Lennox L. Broth (LB) (RPI, powdered or granulated, L24066)Lennox L Agar (RPI, powdered or granulated, L24030)

- Super Optimal Broth with Catabolite Repression (SOC) (SOB with added glucose) (RPI, S25000)
- Antibiotics: Ampicillin (RPI, A40040), Kanamycin (RPI, K22000), Chloramphenicol (RPI, C61000)
- Natural amino acids (see Table 1) (RPI)
- Isopropyl β–d-1-thiogalactopyranoside (IPTG)(RPI, I56000)
- Imidazole (RPI, I52000)
- Sodium Phosphate Monobasic, monohydrate (RPI, S23120)
- Sodium Phosphate Dibasic, heptahydrate (Sigma, S9390)
- Sodium chloride(Fisher, S2713)**.**
- HEPES (Fisher, BP310)
- Tris (Acros, 167620010)
- 3-fluoro-DL-tyrosine (Alfa Aesar, L01479)
- 4-fluoro-DL-phenylalanine (Sigma Aldrich, F5251)
- 5-fluoro-indole (Sigma Aldrich, F9108)
- Ultrapure water 18.2 MΩ x cm

Protein Purification Components

- Lysozyme (Gold Biotech, L040)
- β-mercaptoethanol (Sigma Aldrich, M6250)
- Phenylmethanesulfonylfluoride (PMSF), (RPI, P20270)

NMR Components

- Dimethylsulfoxide (Fisher, D128)
- Ethylene Glycol (Sigma Aldrich, 102466)
- Trifluoroacetic acid (TFA, Sigma Aldrich, T6508)
- Deuterium oxide (Cambridge Isotope Labs, DLM-6)

Prepared Buffer Solutions (Filtered and stored at room temperature)

- Lysis Buffer: 50 mM Phosphate, 300 mM NaCl, pH = 7.4
- Wash Buffer: 50 mM Phosphate, 100 mM NaCl, 30 mM Imidazole, pH = 7.2
- Elution Buffer: 50 mM Phosphate, 100 mM NaCl, 400 mM Imidazole, pH = 7.2
- NMR Buffer for Brd4: 50 mM Tris, 100 mM NaCl, pH = 7.4
- NMR Buffer for KIX: 50 mM HEPES, 100 mM NaCl, pH = 7.2

Protein Characeterization/SDS-PAGE Components:

- Formic Acid (Sigma, 56302)
- Acetonitrile (J.T. Baker, 9853)
- 40% Acrylamide/Bis solution (Bio-Rad, 1610148)
- Ammonium persulfate (Sigma Aldrich, 248614)
- Bis-Tris (RPI, B75000)

- Tetramethylethylenediamine, (TEMED) (RPI, T18000)
- Coomassie Brilliant Blue R-250 (Bio-Rad, 161-0400)

Equipment

- Refrigerated Incubator Shaker
- UV/Vis spectrophotometer (e.g., Varian Cary 50 Bio UV/Vis spectrophotometer)
- High speed centrifuge
- Sonicator (Fisher FB505) with 1/8" microtip (FB 4418)
- FPLC (GE Äktapurifier)
- HiPrep 26/10 Desalting Column (GE Healthcare 17-5087-01)
- 50 mL Superloop (GE Healthcare 18-1113-82)
- Ni-NTA Agarose Beads (Life Technologies R901-01)
- NMR spectrometer (Recommended ^{19}F S/N ≥ 550:1)

Critical: NMR probe must be able to tune to the ^{19}F nucleus (e.g. Bruker BBO probe), For optimal PrOF NMR experiments, a cryoprobe (e.g. Bruker TCI-Prodigy) should be used for increased S/N.)

- ^{19}F tuned NMR probe, e.g. (TCI-Prodigy cryoprobe, Bruker)
- Electrospray Ionization Mass Spectrometer

Critical: Mass spectrometer must be capable of resolving proteins to 2-3 amu for accurate protein characterization as fluorinated proteins will differ by 18 amu per incorporated fluorine.

- SDS PAGE gel apparatus and power supply (for use in protein characterization
- NMR tubes
- Microcentrifuge tube
- 15-ml Falcon Tubes
- 50-ml Falcon Tubes
- Protein concentrators (Millipore Centricon UFC900324, UFC800324)

Software

- NMR processing software (e.g. TopSpin, MNova, etc…)
- Non-linear regression software (e.g. OriginPro, GraphPad, etc…)

4.3.2 Reagent Setup
LB Media/Agar: Autoclave media prior to using.

Defined Media

To simplify the preparation if it is going to be done multiple times, a sterile-filtered stock solution consisting of the reagents listed in Step 3 ($CaCl_2$ through biotin) may be prepared and stored at 4°C for several months; an appropriate volume of stock solution would then be added to the defined media instead of adding the components individually.

1. Combine the amino acids, salts, and nucleotide bases listed below (Table 4.1) in a total volume of 1 L of deionized water, omitting the amino acid that will be labeled, and autoclave the solution. Natural amino acids were all purchased from RPI.

   Pause Point: If the defined media is not going to be used right away, do not continue media preparation beyond this step. Media with just the amino acids and salts with no carbon source (glucose) or vitamins can be stored covered at room temperature for several months.

Table 4.1 Amino Acids, Nucleotide Bases, and Salts

Component	Supplier (Cat. #)	Amount for 1L expression
Alanine	RPI (A20060)	500 mg
Arginine	(A50010)	400 mg
Asparagine	(A50030)	400 mg
Aspartic Acid	(A50060)	400 mg
Cystine HCl	(C81020)	50 mg
Glutamine	(G36040)	400 mg
Glutamic Acid	(G36020)	650 mg
Glycine	(G36050)	550 mg

Histidine	(H75040)	100 mg
Isoleucine	(I54020)	230 mg
Leucine	(L22000)	230 mg
Lysine HCl	(L37040)	420 mg
Methionine	(M22060)	250 mg
Phenylalanine	(P20260)	130 mg
Proline	(P50200)	100 mg
Serine	(S22020)	2.1 g
Threonine	(T21060)	230 mg
Tyrosine	(T68500)	170 mg
Valine	(V42020)	230 mg
Sodium acetate	Macron (7372)	1.5 g
Succinic acid	Sigma Aldrich (S3674)	1.5 g
Ammonium chloride	Fisher (A661)	500 mg
Sodium hydroxide	Alfa Aesar (A16037)	850 mg
Potassium phosphate (dibasic)	RPI (P41300)	10.5 g
Adenine	RPI (A11500)	500 mg
Guanosine	Alfa Aesar (A11328)	650 mg
Thymine	TCI (T0234)	200 mg
Uracil	RPI (U32000)	500 mg
Cytosine	Alfa Aesar (A14731)	200 mg
40% (wt/vol) glucose solution	RPI (G32030)	50 ml
0.01M $FeCl_3$	Sigma Aldrich (157740)	1 ml
1M $MgSO_4$	Sigma Aldrich (M7506)	4 ml

2. Once the solution is sterilized, add the components sequentially, adding $MgSO_4$ last.

Table 4.2 Vitamin Solution

Vitamin Solution		
Component		**Amount for 1-l expression**
Tryptophan	RPI (T60080)	50 mg
$CaCl_2$ dihydrate	Sigma Aldrich (C3306)	2 mg
$ZnSO_4$ heptahydrate	Sigma Aldrich (Z4750)	2 mg
$MnSO_4$ monohydrate	Sigma Aldrich (M7634)	2 mg
Thiamine	RPI (T21020)	50 mg
Niacin	Sigma Aldrich (72340)	50 mg
Biotin	RPI (B40040)	1 mg

3. Adjust the pH to 7.2 using HCl or NaOH. Media can be stored at room temperature for a few weeks, though it is recommended that the glucose and vitamin solutions not be added until immediately prior to use (Table 4.2).
4. Add the necessary fluorinated amino acid(s) or amino acid precursor for appropriate fluorine labeling (Table 4.3).

Table 4.3 Fluorinated amino acids or precursor for defined media

Fluorine Label	**Fluorinated Substitution**	**Amount for 1-L expression**	**Final concentration in media**
3FY	3-fluoro-DL-tyrosine	80 mg	400 μM
4FF	4-fluoro-DL-phenylalanine [a]	29 mg	160 μM
5FW	5-fluoroindole	60 mg	444 μM

[a]Addition of 5 μM of phenylalanine to the defined media when using 4-fluoro-DL-phenylalanine has led to increased protein yield but minimal effects on fluorinated amino acid incorporation.

4.3.3 Equipment Setup

Sonicator Program. Sonication time = 4 minutes at 30% amplitude (this entails eight x 30 second pulses with 60 second rest times between pulses.) Total elapsed time: 12 minutes

Ni Affinity Gradient Elution Method. Set up the FPLC method that will be used in Step 9. In our laboratory we use a GE Äktapurifier with a column packed with Ni-NTA Agarose for affinity purification . The four steps are as follows: Wash the nickel column with 15 column volumes of wash buffer. Perform a gradient elution across 20 column volumes ramping from 0 to 100% elution buffer. Wash with 5 column volumes of 100% elution buffer. Return to 0% elution buffer by applying the reverse gradient in 5 column volumes. UV absorbance is monitored at 280 nm. Total time (at 1 mL/min flow rate): 4 hours

Buffer Exchange Method. Set up the FPLC method that will be used in step 10. In our laboratory, we use a GE Äktapurifier with a GE HiPrep 26/10 Desalting column (17-5087-01). Equilibrate the column with 0.5-1 column volume of buffer prior to loading the sample. Elute the column with 1 column volume of buffer and monitor the UV absorbance at 280 nm. Total time (at 2 mL/min flow rate): 40 minutes.

4.4 **PROCEDURE**

4.4.1 Bacterial Transformation (Timing: 1 day)

1. Prepare bacterial colonies transformed with the plasmid for the protein of interest . using standard transformation and inoculation methods. A detailed sample protocol is provided in Supplementary Method.

Critical As noted in the materials section, choice of cell lines must be made carefully. Auxotrophic cell lines (e.g., DL39(DE3)) are necessary for labeling with 3-fluorotyrosine or 4-fluorophenylalanine. For tryptophan labeling, 5-fluoroindole with non-auxotrophic cell lines (e.g., BL21(DE3)) is recommended for higher protein yields.

Pause Point Culture plates with transformed bacteria may be stored at 4 °C for up to a month.

4.4.2 Fluorinated protein expression Timing: 2 days

Critical: Maintain sterile conditions while working with media. Standard protein expression methods can be used; these steps are those that are used routinely in our laboratory and provide information about when in the process, the fluorinated amino acids should be added.

2. Inoculate primary and secondary cultures using standard *E. coli* expression methods. Detailed steps of a sample protocol are provided in the supporting information.

3. Using a spectrophotometer, determine the optical dispersion at 600 nm (OD_{600}). Remove the culture from the shaker when the OD_{600} is between 0.6 and 0.8. Please note that this value may vary based on the cuvette distance to the detector in the spectrophotometer.

4. Centrifuge the culture for 20 min at 6000 *g* at 4 °C.

5. Decant the LB media and resuspend the pellet in an equivalent amount of the defined media containing the desired fluorinated amino acid or amino acid precursor [see Reagent setup] and appropriate antibiotic.

6. Shake the solution prepared in the previous step at 37 °C and 250 RPM for 90 min as a recovery time for the bacteria then decrease the temperature to 20 °C to cool down the media. Continue to shake for an additional 30 min to allow the solution to equilibrate. Please note that the most suitable temperature and recovery time will vary for each protein. Our lab has generally found the above conditions to work well, though there are cases when a shorter recovery time has produced better results.

7. Induce protein expression by adding IPTG to the solution to a final 1-mM concentration. Continue to shake at 20 °C and 250 RPM for 16–20 h. Please note that the most suitable IPTG concentration and induction time will vary for each protein.

8. Centrifuge the cell culture at 6000 *g* at 4 °C for 20 min. For ease of purification, we recommend centrifuging cultures in 500-ml aliquots or smaller. Decant the supernatant medium and store the cell pellet at –20 °C or –80 °C.

Pause Point The cell pellet may be stored at –20 °C or –80 °C for months.

9. Purify and characterize overexpressed protein. Sample purification methods are described in the supporting information and mass spectrometry characterization is described in **Box 1**. SDS PAGE characterization can be performed at this step to evaluate the presence of the desired protein. However, due to the low resolution nature of SDS PAGE characterization, it will not be able to quantify fluorine incorporation or distinguish between fluorinated variants. (Figure 4.3)

10. Buffer exchange the protein into the desired buffer. In our lab we use a HiPrep desalting column, but other methods work as well (Nap-5 columns, PD-10 columns, and dialysis are all viable options). For PrOF NMR, it is best to avoid buffers with high concentrations of high mobility salts due to their impact on NMR signal sensitivity. When possible, the lower mobility salts (e.g. Tris, HEPES, etc…) are preferred.

Pause Point: The purified protein can be stored at 4 °C or flash-frozen and stored at –20 °C for several months. For long-term storage, we recommend flash-freezing and storing at –20 °C. Exact storage conditions may vary from one protein to another.

Box 1 – Fluorine Incorporation Characterization Timing 0.5-1 h

1. Concentrate protein to at least a low µM concentration. In our lab we utilize centrifugal protein concentrators with either a 3k or 5k MW cutoff. The exact membrane cutoff needed will depend on the size of the protein of interest

2. On an ESI mass spectrometer equipped with a liquid chromatography system, use 0.1% formic acid in water (v/v) and acetonitrile as solvents and a C18 column for separation. Run a gradient elution ramping from 8% acetonitrile to 80% acetonitrile.

3. Select the protein peak on the chromatogram and deconvolute the corresponding mass spectrum.

4. Integrate the peaks corresponding to the protein of interest and its fluorinated variants.

5. Use the following formula to calculate the percent of fluorine incorporation.

$$\%\ Incorporation = \frac{(0\ F\ protein * 0) + (1F\ protein * 1) + \cdots (nF\ protein * n)}{(0F\ protein * n) + (1F\ protein * n) + \cdots (nF\ protein * n)} * 100$$

Where n corresponds to the number of incorporated fluorinated residues.

4.4.3 Protein-Observed Fluorine (PrOF) NMR Timing 5 min–1 h

11. Concentrate the protein solution to 40–50 μM as done in **Box 1**

12. To a microcentrifuge tube, add 2 μl of 0.1% (v/v) TFA, 25 μl of D_2O, and 473 μl of protein solution just prepared. Mix well and transfer solution to a 5-mm NMR tube.

13. Acquire two fluorine NMR spectra. Focus the first experiment on the TFA reference peak, which can be obtained within several scans, and focus the second experiment on the protein. We find that a spectral width (sw) of 10–20 ppm is sufficient with an offset (O1P/tof) of –76.5 ppm for the TFA, –136 ppm for 3FY-labeled proteins, –125 ppm for 5FW-labeled proteins, or –117 ppm for 4FF-labeled proteins. Experiment time and number of scans required will be dependent on availability of a cryoprobe versus a room-temperature probe as well as protein On a 500 MHz NMR with a Prodigy inverse cryoprobe (^{19}F S:N 2100:1) 400 scans are sufficient for proteins KIX and bromodomain Brd4, which would lead to a 4–5-min experiment.

14. Process the data setting the TFA reference peak to –76.5 ppm and applying the same correction factor to the second experiment. Resonance assignments can be performed as described in section Box 2

Box 2: (Optional) Assigning PrOF Resonances Timing ~3 weeks

Note: PrOF NMR screening and binding experiments can still be performed without resonances assigned. The assignments can provide additional structural information but are not necessary to evaluate small molecule binding.

1. Perform site-directed mutagenesis to mutate each amino acid of interest to an alternate amino acid chosen such that it will have very little effect on the protein structure (e.g., Y->F)

2. Obtain PrOF NMR spectra of each mutant (see Steps 11-14 of the Procedure). The resonance that has disappeared corresponds to the mutated amino acid residue.

Box 3: Fragment Mixture Preparation Timing Variable

1. Obtain or prepare concentrated ligand stock solutions (200 mM) in a given solvent (e.g., DMSO or ethylene glycol) for the fragments of interest. Please note that solvent choice will depend on ligand solubility and solvent effects on the protein of interest.

Critical Step The maximum volume of solvent that will be tolerated without significantly perturbing the chemical shifts or the shapes of the ^{19}F NMR resonances must be determined in order to accurately assess ligand binding. We recommend that the final solution is composed of 1–5% organic solvent.

2. Determine the desired number of fragments for each screening mixture. Commonly screening mixtures include 5–10 fragment compounds. Fragment mixtures of five of six compounds will reduce the number of compounds needed in the deconvolution step, if a high hit rate is anticipated.

3. Decide the concentration at which each fragment will be screened. Fragment mixture concentrations will depend on the expected K_d's for ligands in the mixture. Due to the low affinity nature of fragment compounds, mid-μM to low-mM K_ds for individual fragments are common. For example, using six fragments per mixture starting from 200 mM DMSO stock solutions yields mixtures containing ligands at 33.3 mM and a final ligand concentration of 833 μM at 2.5% (v/v) DMSO in the NMR sample. Screening ligands at lower concentrations will yield smaller changes in chemical shift, potentially resulting in more false negatives. If the ligandability and/or druggability of the protein is unknown, a pilot screen with ligands at various concentrations can be performed.

4. Compile the fragment mixtures, taking into account the number of acidic, basic, and neutral compounds present to avoid significant pH dependent effects.

4.4.4 Fragment screening via PrOF NMR Timing Variable

15. Prepare a blank sample (See step 10) with the addition of the selected amount of the organic solvent for the fragment mixtures (no ligand or ligand mixtures), maintaining a total sample volume of 500 µL. (e.g. 2 µL 0.1% (v/v) TFA, 25 µL D_2O, 468 µL protein solution, and 5 µL DMSO)

Critical Step This blank sample must be identical to the screening samples in regards to the protein concentration and amount of solvents used. The only difference should be that the blank contains organic solvent (with no ligands) while the screening samples will contain the same amount of organic solvent (with the ligand mixtures).

16. Prepare the fragment mixture NMR samples in a similar manner to Step 15 with the ligand mixture in place of just the organic solvent.

17. Acquire and process PrOF NMR spectra for each mixture (see Steps 13 and 14)

18. Record the chemical shift of each fluorine resonance and calculate the change in chemical shift for each resonance relative to the 'blank' spectrum.

19. Using the results obtained in the library screen, identify promising mixtures by statistically analyzing the changes in chemical shifts for each resonance from each mixture, and select the mixtures that yield a change in chemical shift between one and two standard deviations above the average change in chemical shifts from all experiments obtained from the library screen.

4.4.5 Deconvolution of fragment mixtures Timing Variable

20. Identify the compounds comprising the mixtures that yielded significant changes in chemical shift.

21. Prepare NMR samples with each of these compounds separately maintaining equivalent ligand concentrations.

22. Collect the PrOF NMR spectra for each new sample (see Step 13)

23. Process and analyze the data as before calculating the changes in chemical shift to identify the ligand or ligands that bind to the protein target. (see Steps 14 and 19)

4.4.6 K_d determination of fragment compounds Timing Variable

24. Select a range of ligand concentrations for the binding isotherm being sure to include points below and above the anticipated K_d.

25. Prepare NMR samples with varying concentrations of ligands, making sure to add solvent until the final volume of ligand solution is equal to the total amount of solvent that was added to the blank prepared in step 15. (1 µL ligand/4 µL DMSO, 2 µL ligand/3 µL DMSO, etc…)

26. Mix the solutions well and transfer them to NMR tubes.

27. Collect PrOF NMR spectra for each sample and analyze data (see Steps 13, 14 and 18).

28. Plot the change in chemical shift (Δδ) as a function of ligand concentration.

29. Fit the data using non-linear regression software (e.g. Originpro or Graph Pad Prism) the following equation to solve for K_d using equation 1 reported above.

$$S = D * \frac{(K_d+L+P)-\sqrt{(K_d+L+P)^2-4PL}}{2P} \quad (1)$$

4.4.7 Timing

Steps 1, Bacterial Transformation: 1 d
Steps 2-10, Fluorinated Protein Expression and Purification: 2 d (plus characterization time; see **Box 1**)
Steps 11-14, Protein Observed Fluorine (PrOF) NMR: 5 min-1 h (plus resonance assignments; see **Box 2**)
Steps 15-19, Fragment Screening via PrOF NMR: Variable (plus fragment mixture preparation time; see **Box 3**)
Steps 20-23, Deconvolution of Fragment Mixtures: Variable
Steps 24-27, K_d determination of Fragment Compounds: Variable
Box 1, Fluorine Incorporation Characterization: 0.5-1 h
Box 2, Assigning PrOF Resonances: ~3 weeks
Box 3, Fragment Mixture Preparation: Variable

4.5 **ANTICIPATED RESULTS**

During fragment screening, resonance assignments facilitate the characterization of the binding site of small molecules and of the environmental changes experienced by the labeled protein side-chain. Fragment mixture hits can result in a large change in chemical shift (e.g., 0.1–1 ppm) due to the responsiveness of fluorine to changes in its chemical environment. The ^{19}F NMR spectra shown in Figure 4.8 show that a peak at 134.5 ppm observed in the spectrum of the fluorinated protein 3FY-KIX undergoes a change in chemical shift of 0.32 ppm in the spectrum of the same protein in the presence of a mixture of different potential ligands.

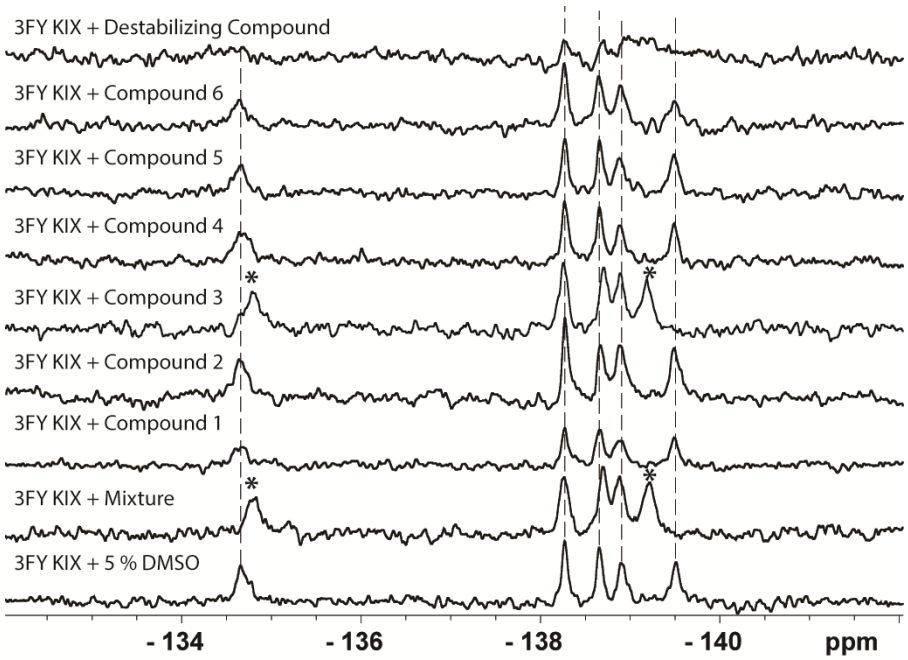


Figure 4.8 Deconvolution of fragment mixture.
The bottom spectrum corresponds to the fluorinated protein with no ligands added. Asterisks denote resonances that have been significantly perturbed. In this particular mixture, compound 3 is responsible for the chemical shift perturbation seen in the mixture. The top spectrum represents a global reduction in signal due to nonspecific effects.

Protein aggregators and denaturants, whose presence is common in many screens, can be detected by global analysis of all of the resonances present in the protein. Global coalescence, broadening, or a decrease in intensity are all indications of non-specific effects induced by the small molecule on the protein, which may lead to false positives in

ligand-observed experiments (e.g., top spectrum in Figure 4.8). Active molecules within a mixture are then identified by testing the molecules from the fragment mixture hit one at a time. Figure 4.8 shows the example of such a 'deconvolution', which helps identifying compound 3 as the active compound, i.e., the compound that caused the change in chemical environment sensed by the fluorine labels in the protein.

Table 4.4 Reproducibility of chemical shifts

Protein/Resonance	Average Chemical Shift (ppm)	Standard Deviation
5FW Brd4 (8 exp)		
W75	-124.813	0.009
W120	-125.464	0.012
W81	-126.358	0.026
3FY KIX (23 exp)		
Y649	-134.668	0.037
Y650	-138.283	0.024
Y658	-138.676	0.025
Y640	-138.894	0.016
Y631	-139.490	0.011

Chemical shifts are easily measured and highly reproducible. In Table 4.4 are reported data from two screens with 3FY-KIX and 5FW-Brd4, showing high reproducibility and low variability of chemical shifts for the protein resonances. In particular, 3FY-KIX, across 23 replicate experiments had a standard deviation in chemical shift over five resonances from 0.011 to 0.037 ppm. The latter measurement was from the broadest resonance, the one assigned to fluorine-labeled Y649. Although only one resonance is observed for fluorine-labeled Y649, the potential for rotamers of the ortho-substituted phenol, from restricted motion in this environment leads to a broadened resonance. In the case of 5FW-labeled Brd4, where the resonances are sharper, the standard deviation range for the fluorine resonances was narrower, from 0.009 to

0.026 ppm. Additionally, although aromatic amino acids are commonly found at protein binding sites,[170] labeled amino acids that are far away from the binding site, serve as important internal control resonances for non-specific binding and other factors that may globally affect the chemical shift. In the 229 compound screen with 5FW-labeled Brd4 the average change in chemical shift (Δδ) of fluorine-labeled W120, which is located outside the ligand binding site, was 0.001 ± 0.014 ppm. However, in the case of W81, which is located at the ligand binding site, the average Δδ of W81's fluorine-labeled analog was 0.033 ± 0.026 ppm (Table 4.5).

Table 4.5 Ligand-induced changes in chemical shift

Protein/Resonance	Average Chemical Shift (ppm)	Standard Deviation
5FW Brd4 (229 exp, 100 μM ligand)		
W120	-125.463	0.014
W81	-126.325	0.069
3FY KIX (111 exp, 833 μM ligand)		
Y640	-138.888	0.011
Y631	-139.485	0.026

Compound affinity is determined for each fragment-hit molecule. Figure 4.9 shows three binding isotherms. The first binding isotherm was used to identify molecule 9B11, a ligand that binds to the protein KIX in the presumed MLL binding site near Y631 (K_d = 1.6 mM, LE = 0.27). The six experiments can be completed in 30 minutes. In the following two binding isotherms, the protein Brd4 was separately labeled with 5FW and 3FY. Titration with Brd4 ligand, acetaminophen, yielded a similar K_d for both proteins (LE =0.44). Titration of two alternately labeled proteins is a useful control to assess any perturbing effects of fluorine on ligand–protein binding. Additionally the magnitude of chemical shift change was substantially larger for the protein labeled with

3FY than for that labeled with 5FW (0.61 ppm vs. 0.14 ppm). These data are consistent with acetaminophen binding farther away from W81 than the affected tyrosine side chains. In some cases, more than one resonance in the NMR spectrum is affected by ligand binding. In these instances, binding isotherms can be obtained based on perturbation of each resonance, thus providing multiple dissociation constant determinations from the same experiment. An example of two affected resonance perturbations for the 5FW-labled bromomdomain, BrdT, in the presence of a new fragment is shown in Figure 4.6 and Figure 4.10, yielding comparable dissociation constants of 250 μM and 210 μM. The larger magnitude of the change in chemical shift is consistent with ligand binding near the presumed WPF resonance W50.

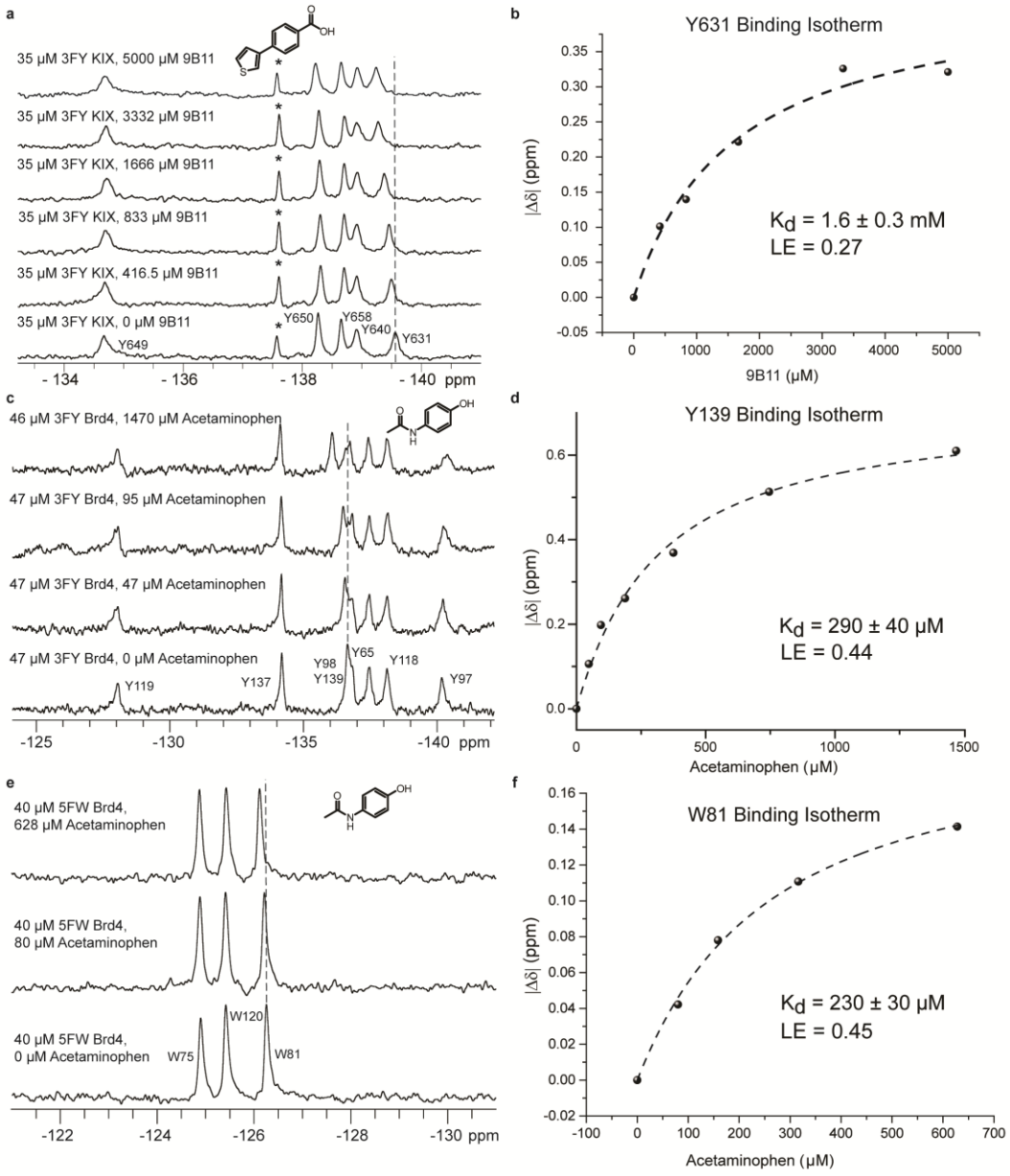


Figure 4.9 Binding isotherms from PrOF NMR titrations.
(a) Stacked spectra for the titration of 3FY KIX with molecule 9B11 (see structure at the top of the panel). A dashed line is added for reference. The asterisk denotes a partial degradation resonance. (b) Binding isotherm generated by monitoring the change in chemical shift of fluorine-labeled Y631 as a function of small-molecule concentration. (c) Stacked spectra for the titration of 3FY Brd4 with acetaminophen (see structure at the top of the panel). A dashed line is added for reference. (d) Binding isotherm generated by monitoring the change in chemical shift of fluorine-labeled Y139 as a function of small-molecule concentration. (e) Stacked spectra for the titration of 5FW Brd4 with acetaminophen (see structure at the top of the panel). A dashed line is added for reference. (f) Binding isotherm generated by monitoring the change in chemical shift of fluorine-labeled W81 as a function of small-molecule concentration.

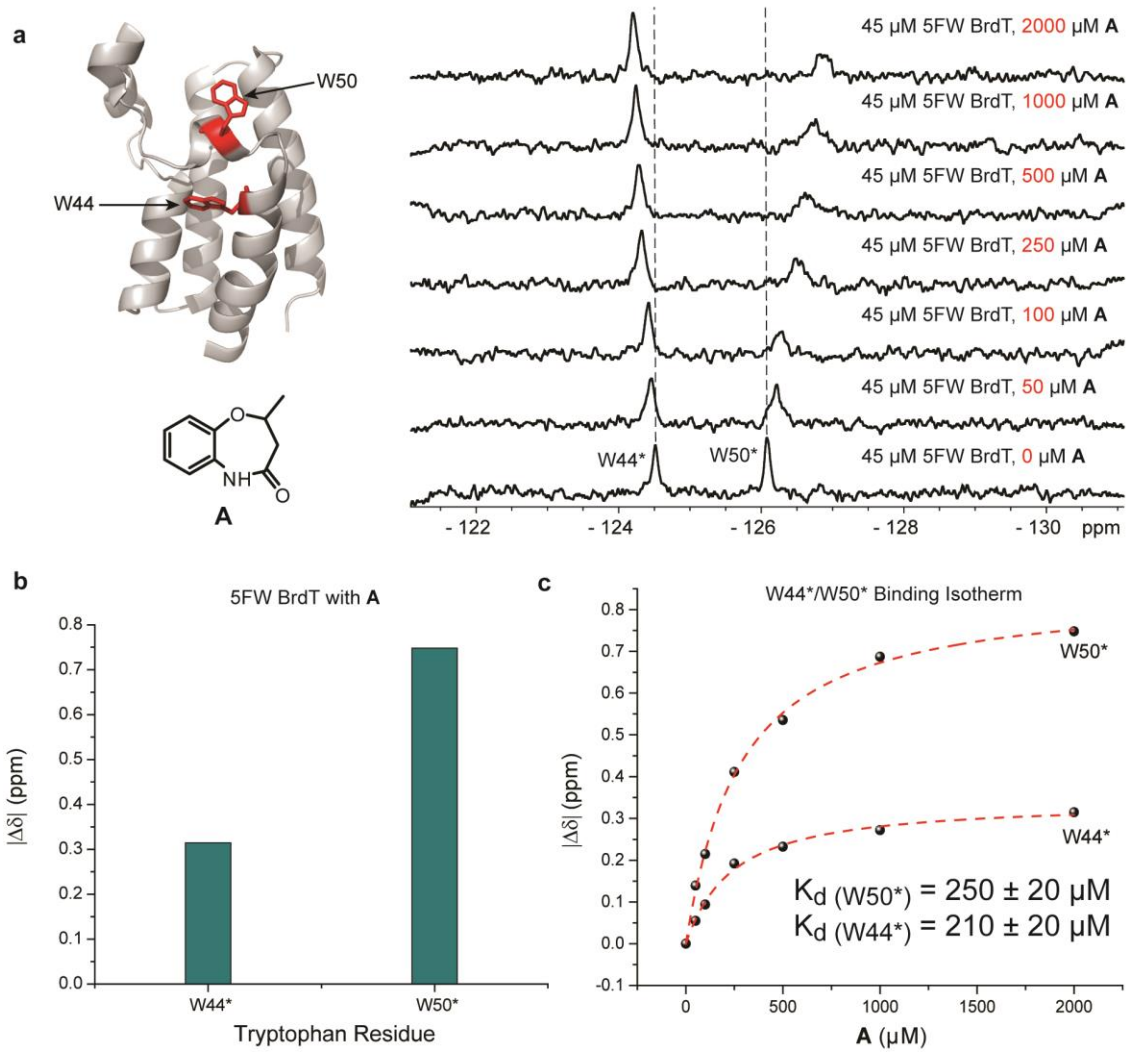


Figure 4.10 Titration of A with BrdT.
(a) Left: Ribbon diagram of BrdT (PDB Code: 4FLP) with tryptophan side chains indicated as sticks. Right: Stacked PrOF NMR spectra with increasing concentration of A. (b) Absolute value of chemical shift perturbations for all 5FW BrdT tyrosine resonances at 2 mM A. (c) Binding isotherm of both W44 and W50 perturbations for the titration with A.

Hajduk et al. have demonstrated the efficacy of NMR-based screening methods to assess the druggability of a protein target.[175] Proteins with screening hit rates that are lower than 0.1% using fragment libraries are characterized to have low druggability. Upon completion of the fragment screen, PrOF NMR can also provide druggability information based on the screening hit rate, both for different proteins or different binding sites within a single protein. As an example, of the 508 compounds screened

against KIX, the MLL site was preferential for ligand binding (0.8% hit rate) over the CREB site (0% hit rate),[193] consistent with previously reported work.[186]

While we have demonstrated the high speed of PrOF NMR experiments for screening,[193] it is dependent on the available equipment. Without a cryoprobe, the same experiment will take 16-20 times longer, even on a magnet of the same field strength. To mitigate this time challenge in the absence of higher end equipment, as may be the case at a primarily undergraduate institution (PUI), we sought to develop solution additives that could shorten experiment time.

As discussed in Chapter 1, paramagnetic metals have been well established in the literature as relaxation agents in the magnetic resonance field.[123] We sought to take advantage of this property of paramagnetic metals with the goal of being able to apply it to PrOF NMR.[107] In the presence of paramagnetic metals, the relaxation times of the fluorine nuclei on the side chains of proteins would be shorten, thereby allowing us to shorten the overall recycle delay for the NMR experiment, leading to an ultimately shorter experiment (Figure 4.11).

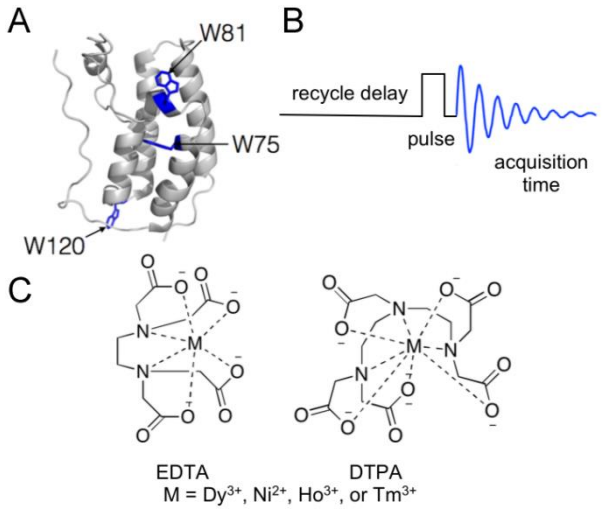


Figure 4.11 Overall schematic for PRE additives for PrOF NMR.
A) Brd4 with sites of fluorine-labelled tryptophans shown as sticks. B) Standard 1D ^{19}F NMR pulse sequence with a 90° pulse. C) Structures of EDTA and DTPA chelates used in these experiments.

To find a metal that would shorten T_1 while minimally affecting T_2, the effects of several metals on ^{19}F T_1 and T_2 times were compared. To minimize interactions between the metal and protein or small molecule, the metals were chelated, initially with the common chelator EDTA. For subsequent work we moved to DTPA, which forms a more stable metal chelate (Figure 4.11C).[236-237] We initially tested nickel, dysprosium, holmium, and thulium chelates to determine their effect on the relaxation properties of small molecule 4-fluorophenylalanine (4FF).[171, 194]

In addition to changing the sample's relaxation properties, the three lanthanides tested also cause lanthanide-induced shifts in the NMR spectrum, which have been utilized in 2D ^{1}H–^{1}H and ^{1}H–^{15}N systems to resolve resonance overlap.[237] Because of the environmental sensitivity of the ^{19}F nucleus, most PrOF NMR spectra of folded proteins show strong resonance dispersion and would not significantly benefit from lanthanide-induced shift reagents. Chelated Ni(II) showed the strongest effect on T1 of the metals tested (Figure 4.12A), was the most economical, and did not introduce the additional complications of a shift reagent. Therefore, it was selected for further study.

To determine the optimal chelated Ni(II) concentration for protein studies, we explored the effect of a range of concentrations of chelated Ni(II) on the relaxation properties of 4FF (Figure 4.12A) as well as 5-fluoroindole (Figure 4.12B), the side chain present in proteins that have been ^{19}F-labelled at their tryptophan residues. The T_1 values of both of these molecules at Ni(II) concentrations up to 320 mM for 4FF and 200 mM for 5-fluoroindole were determined by inversion recovery experiments. T_2* values were determined from the peakwidth at half height. Because of the importance of peakwidth for resolving spectral resonances, T_2* was used rather than T_2. For both 4FF and 5-fluoroindole, the highest Ni(II) concentration tested led to a more than a 100-fold decrease in T1 values (Table 4.6). T_2* also decreased with increasing chelated Ni(II) concentration, leading to approximately a 5-fold increase in peak width at 200 mM chelated Ni(II). Concentrations of 20–50 mM provide an optimal balance between T1 reduction and T_2* reduction and agree well with the 50 mM Ni(II) used by Cai et al. in ^{1}H, ^{15}N, and ^{13}C experiments.[70]

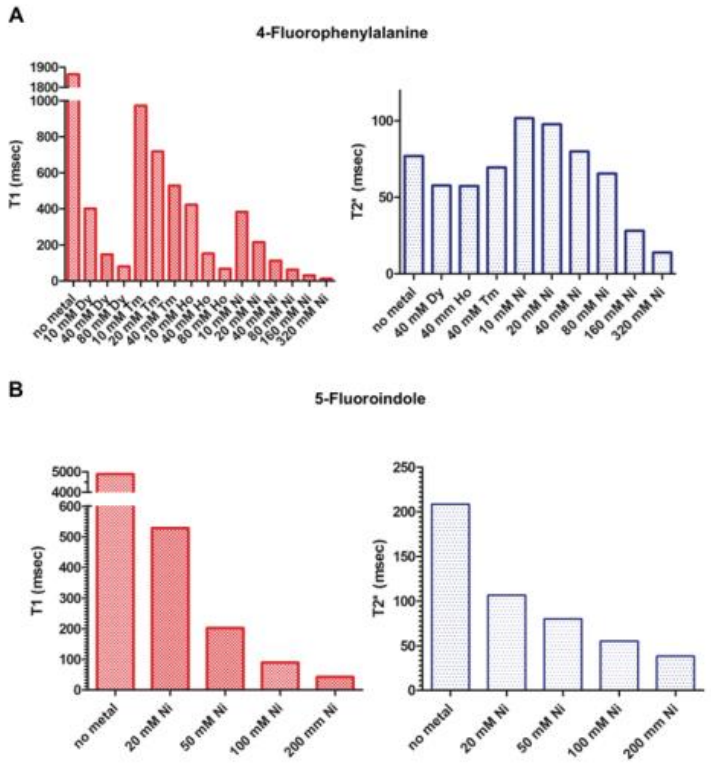


Figure 4.12 Optimization of paramagnetic metal additives.
A) T_1 and T_2* measurements on 4-fluorophenylalanine upon the addition of Dy(III), Ho(III), Ni(II), and Tm(III) as EDTA chelates. All metals cause decreases in T_1 but Ni(II) causes the largest and most rapid decreases for increasing Ni(II) concentrations. B) T_1 and T_2* measurements on 5-fluoroindole with increasing concentrations of Ni-DTPA

Table 4.6 T_1 and T_2* values for fluorine resonances of small molecules

Compound	Ni-DTPA (mM)	T_1 (ms)	T_2* (ms)
5-fluoroindole	0	4890	208
	20	528	107
	50	202	80
4FF	0	1870	77
	20	216	97
	40	113	80

We then moved to testing the effects of Ni-DTPA on the model protein Brd4. We were able to demonstrate that under optimized conditions (20 mM-Ni-DTPA, 0.24 sec recycle delay), we could obtain spectra of comparable quality in 2 minutes compared to our standard 5 minute experiment (Figure 4.13). Protein stability and protein-ligand

interactions are unaffected by paramagnetic additives, making this addition useful for increasing throughput. Additional work is ongoing for applying this technique at a PUI where they are more limited on instrumentation and can therefore benefit significantly from reduced experimental time.

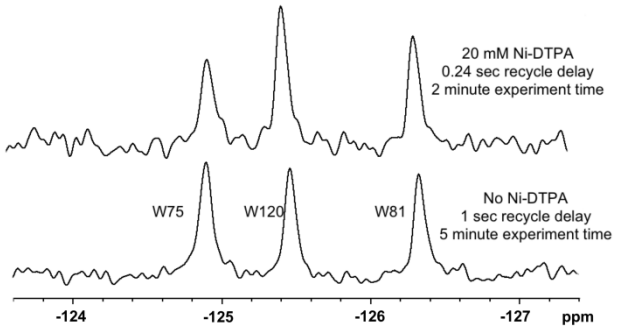


Figure 4.13 Comparison of ^{19} NMR spectra of 50 µM 5FW Brd4 in the presence and absence of paramagnetic additive.
The addition of 20 mM Ni-DTPA allows a 60% reduction in experiment time while preserving the signal-t0-noise ratio of W81, perturbations of which report on ligand binding. The signal-to-noise ratio of W80 without Ni-DTPA was 11.8 and with Ni-DTPA was 10.9

In conclusion, PrOF NMR offers the medicinal chemist a useful screening platform for small molecule fragments in early stage ligand discovery campaigns either as a primary or secondary screen informing future ligand optimization efforts. Since the seminal studies by Sykes et al. on alkaline phosphatase,[137] ^{19}F probe developments and improvements in fluorinated amino acid incorporation have led to ligand identification and affinity quantification with small-to-medium-sized proteins in a time-efficient fashion.[194] However, in all cases, fluorine perturbation to protein function should be assessed. While current limitations to PrOF NMR exist, including mixture deconvolution steps and CSA relaxation effects on fluorine nuclei located on protein side chains with limited mobility or in large proteins. These challenges are beginning to be addressed by complementary ligand-observed NMR methods as well as new fluorinated amino acid labeling strategies.[74]

4.6 **TROUBLESHOOTING**

Table 4.7 Troubleshooting Table

Step	Problem	Possible Reason	Solution
9	Low protein yield relative to protein expression yields using natural amino acids	Cell line may not be robust for protein expression	Change cell lines. Auxotrophic cell lines tend to provide lower yields than non-auxotrophic cell lines
			Use glyphosate to induce auxotrophy in a standard cell line [e.g. BL21(DE3)][199, 238]
		Poor compatibility with tRNA synthetase	Cotransform with a plasmid to increase tRNA synthetase levels[233]
			Utilize cell-free methods[239]
		Toxic levels of fluorinated amino acids	Reduce the concentration of fluorinated amino acids
Box 1	Poor label incorporation (< 90%)	Non-optimized recovery time	Experiment with longer recovery times following the media swap
		Not enough fluorinated amino acid or amino acid precursor	Increase the amount of the fluorinated amino acid

13	Extremely broad NMR resonances	Magnetic field strength is too high or side chain dynamics are too slow, which can cause significant broadening of resonances due to chemical shift anisotropy Protein is too large	We find ideal field strengths to be 500 MHz – 600 MHz. The signal sensitivity and resolution benefits of higher field strength magnets can be cancelled out by CSA and shortened T_2 relaxation. Use alternate labeling strategies that utilize fluorinated amino acids with more rapidly rotating fluorine groups (e.g., trifluoromethyl)[230, 232]
	Unknown fluorine resonances in the NMR spectrum	A fluorinated impurity is present	Fluorinated impurities in your sample may be removed by a buffer exchange Alter spectral window. Polymer background is most pronounced beyond -140

14	Fluorine referencing problems	Fluorine reference is sensitive to solution conditions or the presence of small molecule ligands	Alternative fluorine references (e.g. trifluoroethanol) can be used Alternative NMR pulses (e.g. ERETIC) can be utilized.[240-241]
Box 2, Step 2	Site directed mutagenesis failure	The specific amino acid may be necessary for the protein to adopt its proper secondary structure or affects more than one resonance	Site-selective labeling of 4FF and 3FY can be used to label individual residue positions rather than globally changing all of a given amino acid[100, 242] Kitevski-LeBlanc et al. describe a multidimensional NMR method to assign resonances[101] Paramagnetic metals (e.g., Gd^{3+}) can be used to broaden surface exposed residues
19	Protein signal overwhelmed by additional fluorine resonances	A fluorine-containing molecule is present in the mixture	Deconvolute the mixture to ascertain if other ligands are binding Test fluorinated molecules separately and at lower concentrations Use a different fluorinated amino acid labeled protein
23	No deconvoluted compounds exhibit the same change in chemical shift as the mixture	Small molecules may have additive effects on the protein signals	Select the compound that contributes the most towards the change Prioritize other mixtures with more distinct effects from the compounds
29	Data do not fit well for K_d determination (Poor R^2 value)	Saturation point not yet reached, making it difficult to accurately determine a K_d using the equation Nonspecific or multi-site binding	Increase the concentration of the ligand to approach saturation Deprioritize this compound Change non-linear regression analysis to include multiple binding sites

Chapter 5. Development of selective small-molecule inhibitors for the bromodomain BrdT for male contraception.

A portion of this chapter was reproduced with permission from "BET Bromodomain Inhibitors with One-Step Synthesis Discovered from Virtual Screen," A.M. Ayoub, L.M.L. Hawk, R.J. Herzig, J. Jiang, A.J. Wisniewski, C.T. Gee, P. Zhao, J. Zhu, N. Berndt, N.K. Offei-Addo, T.G. Scott, J. Qi, J.E. Bradner, T.R. Ward, E. Schonbrunn, G.I. Georg, and W.C.K. Pomerantz. *Journal of Medicinal Chemistry,* **2017**, *60* (12), 4805-4817.

Note that this work was highly collaborative with several members of the Georg lab. In particular some of this work will also be published in the dissertation of Andrea Wisniewski. All fluorescence polarization experiments performed here were done by Alex Ayoub from the Pomerantz lab. Alpha-screen competition data was provided by our collaborators at the Broad Institute.

Motivation: This project was a collaborative effort between the Pomerantz lab in Department of Chemistry and the Georg lab in the Department of Medicinal Chemistry. As part of the cross-disciplinary training under the NIH-funded Chemical Biology Interface Training Grant, the idea was to gain experience in synthetic medicinal chemistry to complement the expertise that had been developed in biophysical assays. This chapter will encompass my contribution to three different projects, all related to the development of BrdT inhibitors. In the first project, we pursued a ligand deconstruction project with the goal of developing a selective BrdT ligand. In the second and third projects, we sought to use PrOF NMR as a secondary screen and to help guide the SAR and ligand development process.

5.1 **Introduction**

5.1.1 An overview of epigenetics

The field of epigenetics has gained a significant amount of traction in recent years for its biological relevance in a cancers,[243] cardiovascular diseases,[244] inflammation[245] and other diseases.[246] While the field of genetics addresses questions about genomic DNA, the field of epigenetics and the "histone code hypothesis" address questions about chromatin accessibility and the process by which certain portions of DNA are transcribed and others are not. Epigenetics, as noted by Feinberg and others in the field, is "the information heritable during cell division other than the DNA sequence itself."[246] Even though every cell in the human body contains the same nuclear DNA sequence, other factors, such as environmental stimuli and nutrition, can contribute to epigenetic modifications leading to different cellular outcomes and gene expression.[245]

The accessibility or inaccessibility of DNA for transcription, therefore, is an important and necessary factor to consider. In the nucleus of a cell, DNA is tightly packed and organized in nucleosome subunits around octamers of histone proteins. This octamer is formed from four different histone partners: H2A, H2B, H3, and H4. H3 and H4 form a tetramer while H2A and H2B form two dimers. Together, they form an octamer which serves as a scaffold around which two superhelical turns of DNA can wrap. Strands of nucleosomes, then, form chromatin, which exists in two forms, euchromatin which is accessible for transcription, and heterochromatin, which is tightly packed and inaccessible. The histone code hypothesis posits the idea that the control of whether DNA is in a euchromatic or heterochromatic state is dependent on posttranslational modifications on the amino acid side chains or "tails" of histone proteins.[247]

Histones can be post translationally modified to effect various biological functions.[248-249] The proteins that install, recognize, and remove these modifications are colloquially referred to as "writers", "readers" and "erasers." The "writer" proteins (e.g., acetylases, methylases, and phosphorylases) will add a given post-translational modification (PTM) to a histone, thereby marking it for a particular biological effect. The "reader" proteins (e.g., bromodomains and chromodomains) recognize the PTMs

installed by the "writers" to carry out a given biological function. Once the task has been completed, the "eraser" proteins (e.g., deacetylases, demethylases, and phosphatases,) will remove the PTM. Because these proteins possess such a vital role in biological activity, they are often found to be dysregulated in disease states, making them attractive therapeutic targets.[246]

5.1.2 BET Bromodomains as therapeutic targets

Bromodomains are a class of epigenetic "reader" proteins that recognize acetylated lysine residues on histone tails. There are 61 known human bromodomains that are broken up into eight different families based on structural similarity.[250] The second class of bromodomains are known as the Bromodomain and Extra-Terminal (BET) family. The BET family consists of four different bromodomain-containing proteins, Brd2, Brd3, Brd4, and BrdT, with each having a set of tandem N-terminal bromodomains, commonly referred to as domain 1 and domain 2 (Figure 5.1).

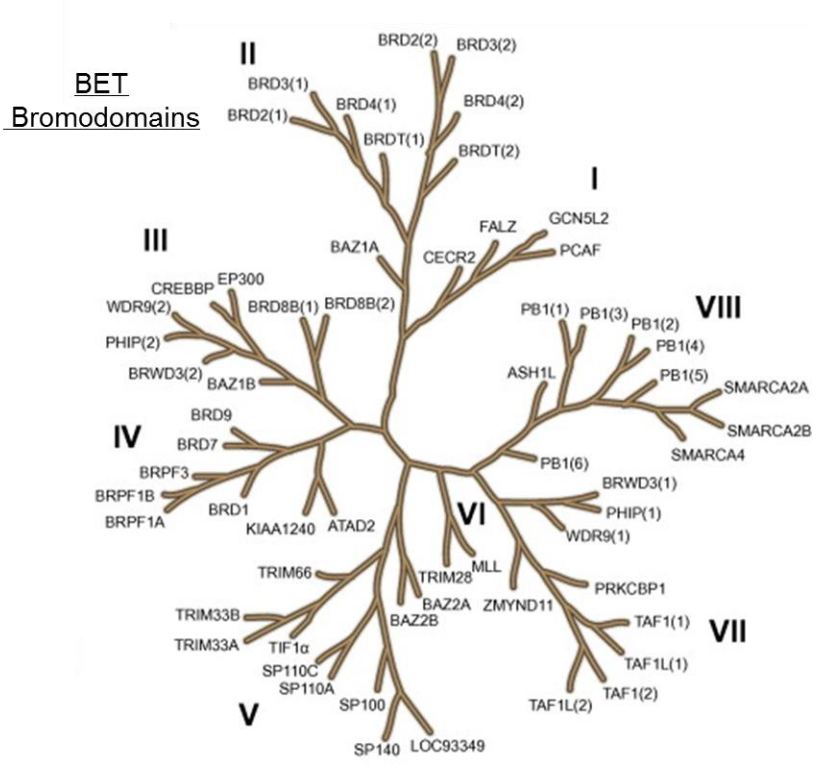


Figure 5.1 Phylogenetic tree of the human bromodomains.
The BET family (Class II) is shown at the top. Adapted from Filippakopoulous et al.[251]

BET bromodomains received tremendous attention in 2010 when Bradner and coworkers reported on the small-molecule JQ1 as a selective BET bromodomain inhibitor.[189] In their seminal paper, Filippakopoulos et al. reported a 50 nM and 90 nM affinity of JQ1 for Brd4 (D1) and (D2) respectively. They were also able to demonstrate the ability of JQ1 to modify cancerous cells to differentiate into normal non-cancerous cells (Figure 5.2). At the same time, researchers from GlaxoSmithKline also reported a selective BET bromodomain inhibitor, I-BET.[252] Since then, many BET bromodomain inhibitors have been discovered and reported (

Table 5.1).

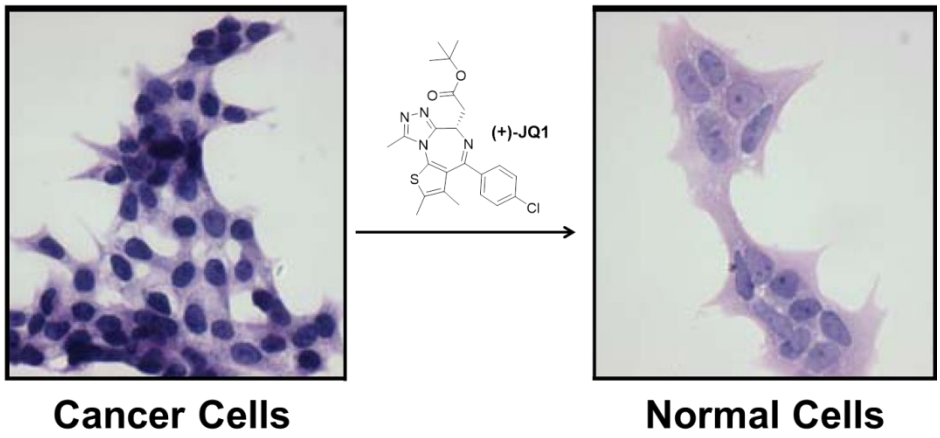
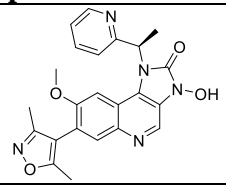
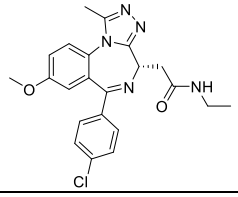
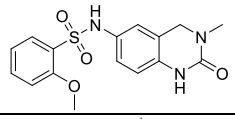
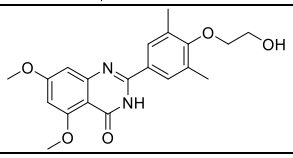
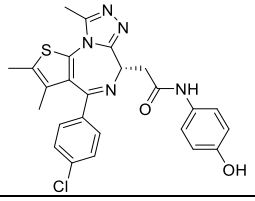


Figure 5.2 Differentiation of cancerous cells into normal cells after treatment with (+)-JQ1.
Reprinted by permission from Macmillan Publishers Ltd: Nature, Filippakopoulous et al[189] 2010.

Table 5.1 Select examples of BET bromodomain inhibitors

Compound		Citation
I-BET151		Seal et al.[253]
I-BET762		Mirguet et al.[168]

PFI-1		Picaud et al[254]
RVX-208		Picaud et al[255]
OTX015		Coudé et al.[256]
A list of the 18 BET bromodomain inhibitors currently in clinical trials has also been compiled by von Schaper.[257]		

5.1.3 BrdT as a target for male contraception

BrdT, or the Bromodomain Testis-Specific, as its name suggests, is localized to the testes and is known to be involved in spermatogenesis.[258] It has been shown that BrdT knockout mice are sterile but still viable.[259-260] Additionally, male mice that were treated with the pan-BET inhibitor JQ1 exhibited sterility while being dosed with the compound but regained normal biological function once treatment with JQ1 was stopped (Figure 5.3).[261] This discovery validated BrdT as a viable target for male contraception.

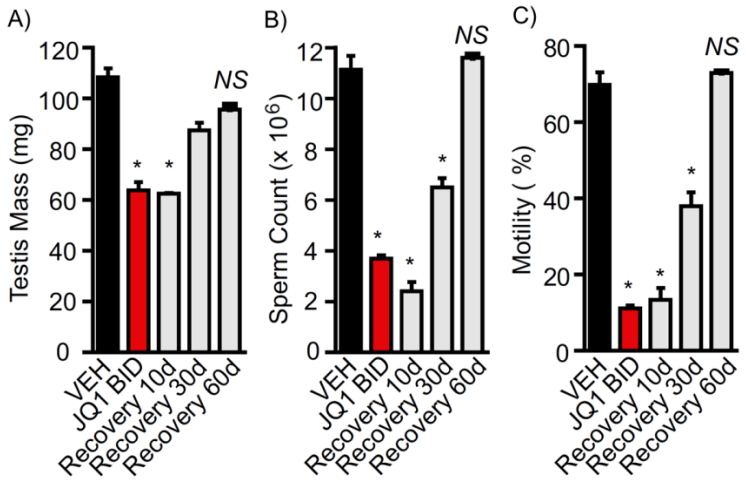


Figure 5.3 Dose-responsive effect of (+)-JQ1.
A) Testis mass, B) Sperm Count, and C) Sperm motility. Adapted from Matzuk et al.[261]

5.1.4 Selectivity challenges for bromodomain inhibitors

Because bromodomains recognize acetylated lysine residues on histone tails, they are often very similar in structure.[189] Conserved regions, like the WPF shelf, are present in a large percentage of bromodomains, making it difficult to develop selective inhibitors. Selectivity among bromodomains that are in the same class is an even harder challenge. Inhibitors like (+)-JQ1 and I-BET762 are referred to as pan-BET inhibitors because they selectively target the BET family but are not selective for specific proteins within the BET family.[168, 189] Therefore, developing ligands that are selective not only for a single class of bromodomains but also for a specific bromodomain within a given class is still a difficult challenge.

Given the amount of research that has now gone into the BET family, the most well studied of the eight bromodomain families, it is difficult to find unique chemical space, making alternative screening techniques, ligand deconstruction, and additional fragment screening campaigns all important areas of research to explore further. New techniques for small-molecule screening can help identify new scaffolds to explore, especially if that technique has some way to also identify selective ligands from the outset. Ligand deconstruction studies can provide insight into specific binding modes of ligands. By deconstructing a compound, it is possible to identify key moieties for binding and potentially aid in determining how the ligand can be made selective for one bromodomain over another. Additional fragment screening, allows for the discovery and exploration of new core scaffolds that may differ from existing core scaffolds (e.g., benzodiazepines,[262] tri-substituted imidazoles,[99] and pyridones[263])

Recently, Urick et al. reported on the first ligand selective for BPTF, a non-BET bromodomain through a dual-protein PrOF NMR screen.[99] By combining both fluorine-labeled Brd4 and fluorine-labeled BPTF in the same NMR tube, they were able to identify ligands from the published kinase set (PKIS I and II) that selectively perturbed Brd4 resonances, ligands that selectively perturbed BPTF resonances, and ligands that bound to both proteins. This dual-screening approach led to the discovery of an aryl-urea-containing compound AU1, the first selective BPTF ligand, as well as additional Brd4-selective ligands with a tri-substituted imidazole core (Figure 5.4). AU1 was also shown

to inhibit BPTF transcription in cells. This study serves as just one example of how novel screening techniques can help to address this selectivity challenge in bromodomains.

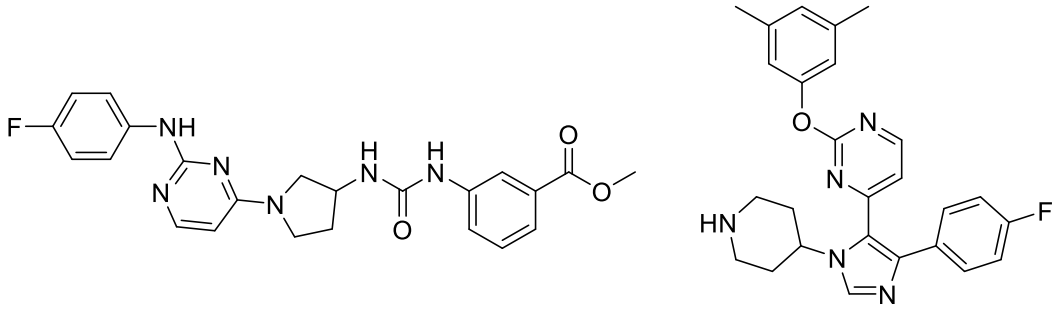


Figure 5.4 Ligands identified from a dual screen of Brd4 and BPTF.
AU1 is selective for BPTF while SB-284851-BT is selective for Brd4.

5.2 **Fluorine-Labeling of BrdT for PrOF NMR**

As established by Mishra and Urick et al., fluorine-labeling of bromodomains allows for the use of protein-observed fluorine NMR spectroscopy (PrOF NMR) which is a useful technique for studying ligand binding in bromodomains.[98] In particular, labeling a bromodomain with a 5-fluorotryptophan (5FW) is particularly efficacious due to the conserved tryptophan in the WPF shelf adjacent to the acetylated lysine binding pocket (Figure 5.5A). This tryptophan is well positioned to report on ligand binding occurring within the pocket. BrdT contains two tryptophans (W44 and W50) which have been previously fluorine-labeled for PrOF NMR.[98] W50 is found in the WPF shelf and is the more upfield resonance while W44 is found close to but outside of the binding site and is the more downfield resonance (Figure 5.5B).[166] BrdT also has seven tyrosine residues that could be fluorine labeled for PrOF NMR. For our studies, we chose to label BrdT with 5FW (5FW BrdT) due to ease of expression, high yields, and high purity.

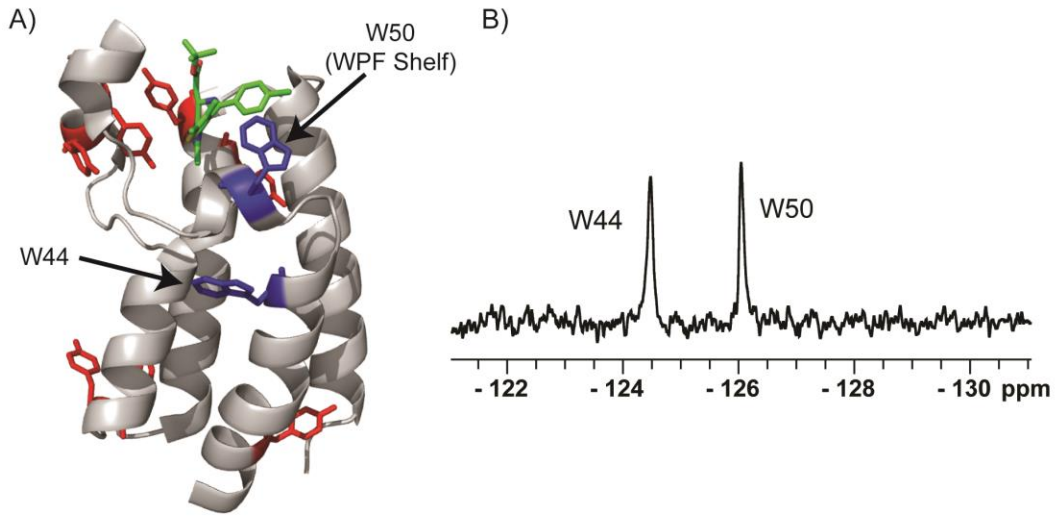


Figure 5.5 BrdT for PrOF NMR.
A) Crystal structure of BrdT with (+)-JQ1. Tryptophans are shown as sticks and designated with arrows. W50 is noted to be in the WPF shelf. Tyrosines are also shown as sticks and highlighted for reference. (PDB ID: 4FLP) B) Sample PrOF NMR spectrum of 5FW BrdT.

5.3 **Ligand Deconstruction of BI-2536 for BrdT**

5.3.1 BI-2536 bromodomain activity

In the process of discovering new bromodomain inhibitors, it was reported that a number of kinase inhibitors inhibited bromodomains. One of these compounds, as reported independently by both the Knapp and Schonbrunn labs, was the compound BI-2536, a PLK1 inhibitor with a pteridinone scaffold at the core (Figure 5.6).[264-265] . BI-2536 has a 25 nM IC_{50} for Brd4 by an Alpha Screen competition assay[264] and a 37 nM K_d by ITC.[265] It also possesses about a ten-fold preference for Brd4 over BrdT with its IC_{50} for BrdT being determined to be 260 nM by Alpha Screen.[264]

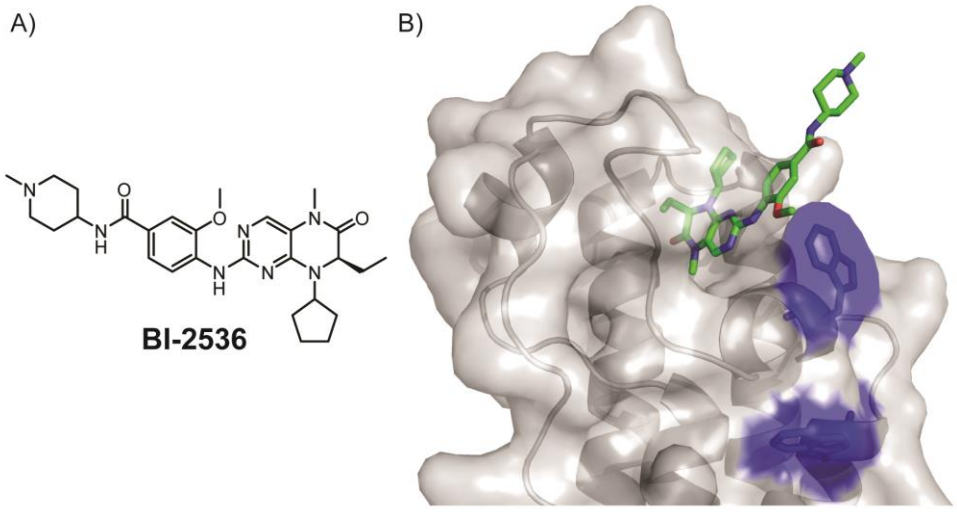


Figure 5.6 Crystal structure of Brd4 crystallized with BI-2536.
A surface representation of the protein is shown with the small molecule represented as sticks. Tryptophans are highlighted to show proximity to binding site. (PDB ID: 4OGI)

A structure activity relationship (SAR) study of BI-2536 against Brd4 was recently reported by Chen et al.[266] In their manuscript, they analyzed a wide range of substitutions and compared the affinity of their new analogs for Brd4 and PLK1 with the affinity of BI-2536. In most cases, there was either a decrease or a modest change in affinity. They were able to identify one change (a replacement of the cyclopentyl group at the N-8 with a 3-bromobenzyl group) that resulted in a seven-fold improvement in affinity. Their other key discovery in the manuscript was their ability to reverse the selectivity and generate compounds that had a better affinity for Brd4 than for PLK1. This study demonstrated that it is possible to take a drug compound with polypharmacological activity and modify it to make it more selective for a desired target (Figure 5.7).

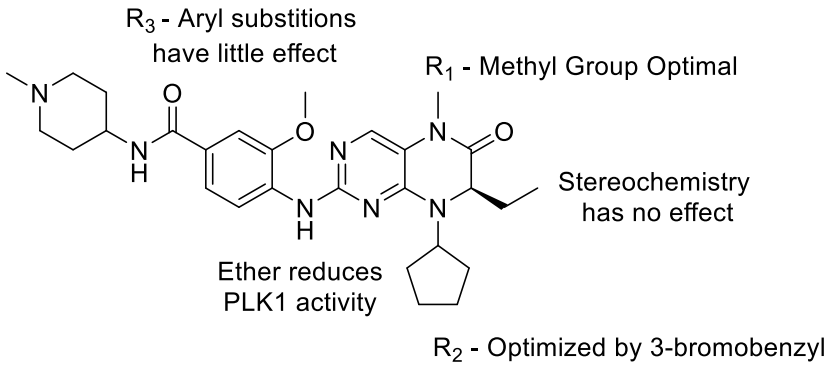


Figure 5.7 Summary of results from Chen et al. SAR Study with BI-2536.

5.3.2 Initial deconstruction efforts

To date, there has not been a ligand deconstruction study performed for BI-2536 against BrdT. We sought to fill this gap in knowledge through synthesizing a range of deconstructed analogs of BI-2536 with the goal of developing a compound that would be selective for BrdT over Brd4. Other Georg lab members synthesized the original set of deconstructed analogs. In all cases, the 7-ethyl-pteridin-6-one was maintained with R_1 at the N5 position, R_2 at the N8 position, and R_3 at the 2 position (Figure 5.8). One series of compounds explored different ring substitutions at the R_2 with a hydrogen in the R_1 and a chloro group at R_3. Another series explored substitutions at both the R_1 and R_2 positions while maintaining a chloro substitution at R_3. A third series then primarily explored the R_3 substitutions with the R_1 remaining as a methyl group and the R_2 either substituted as a cyclopentyl or a cyclohexyl substitution. These compounds were all analyzed by differential scanning fluorimetry (DSF), and those results will be reported in the dissertation of Andrea Wisniewski.

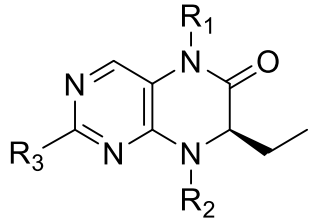
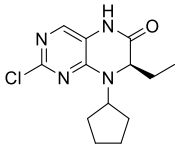
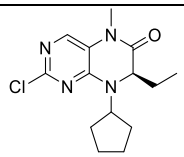
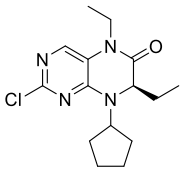
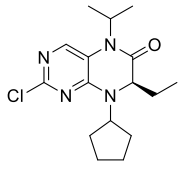
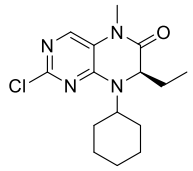
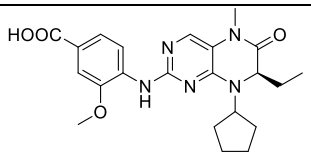
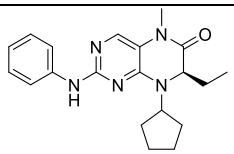
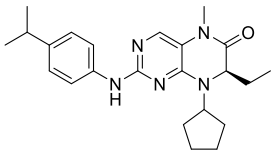
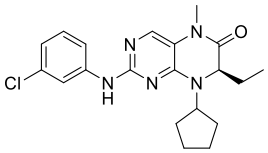
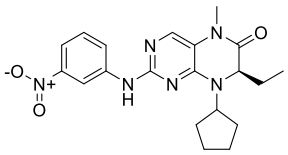


Figure 5.8 Core scaffold of BI-2536 deconstructed analogs.

My initial contribution to this project was to analyze a subset of these compounds that were synthesized by Georg lab members (**G1-G10)** by PrOF NMR (Table 5.2). With the compounds that were tested, one of the interesting trends that was observed was as the R_1 alkyl substitution grew larger (**G2, G3, G4**), the affinity decreased, identifying the methyl group as the substitution that provided the highest affinity (Figure 5.9, Figure 5.10, Figure 5.11, Figure 5.12). This observation is consistent if one extrapolates a similar binding pose of BI-2536 with BrdT as with Brd4 (Figure 5.6). The R_1 substitution is positioned into the binding pocket, so as larger substitutions are placed at this position, they may experience greater effects from steric hindrance. Three of the compounds that were tested demonstrated low solubility and were therefore removed from further consideration. Another appeared to be a weakly-binding ligand and was also removed from further consideration. However, **G6** demonstrated slow-exchange kinetics upon binding, making it our lead compound from the initial set of deconstructed analogs (Figure 5.13).

Table 5.2 Summary of BrdT data from initial set of deconstructed analogs

Compound		K_d by PrOF NMR (µM)	LE	LLE	IC_{50} (FP) (µM)	K_i (FP) (µM)
G1		ND				
G2		12.6 ± 3.1 29.5 ± 4.4	0.332 0.307	2.28 1.91	13	1.2
G3		23.4 ± 7.1 111.0 ± 63.9	0.299 0.255	1.67 0.995	92	17
G4		37.9 ± 3.0 90.5 ± 17.0	0.272 0.249	1.14 0.763	190	38

G5		48.5 ± 8.3 117.2 ± 44.8	0.278 0.253	1.27 0.891	52	9.2
G6		***			0.4305	< 0.330
G7		ND				
G8		ND				
G9		ND				
G10		ND				

K_d values were measured using both W44 (top) and W50 (bottom) and 44 μM 5FW BrdT
ND - indicates compounds that did not bind, were insoluble, or bound very weakly
*** indicates compounds that bound with slow exchange kinetics

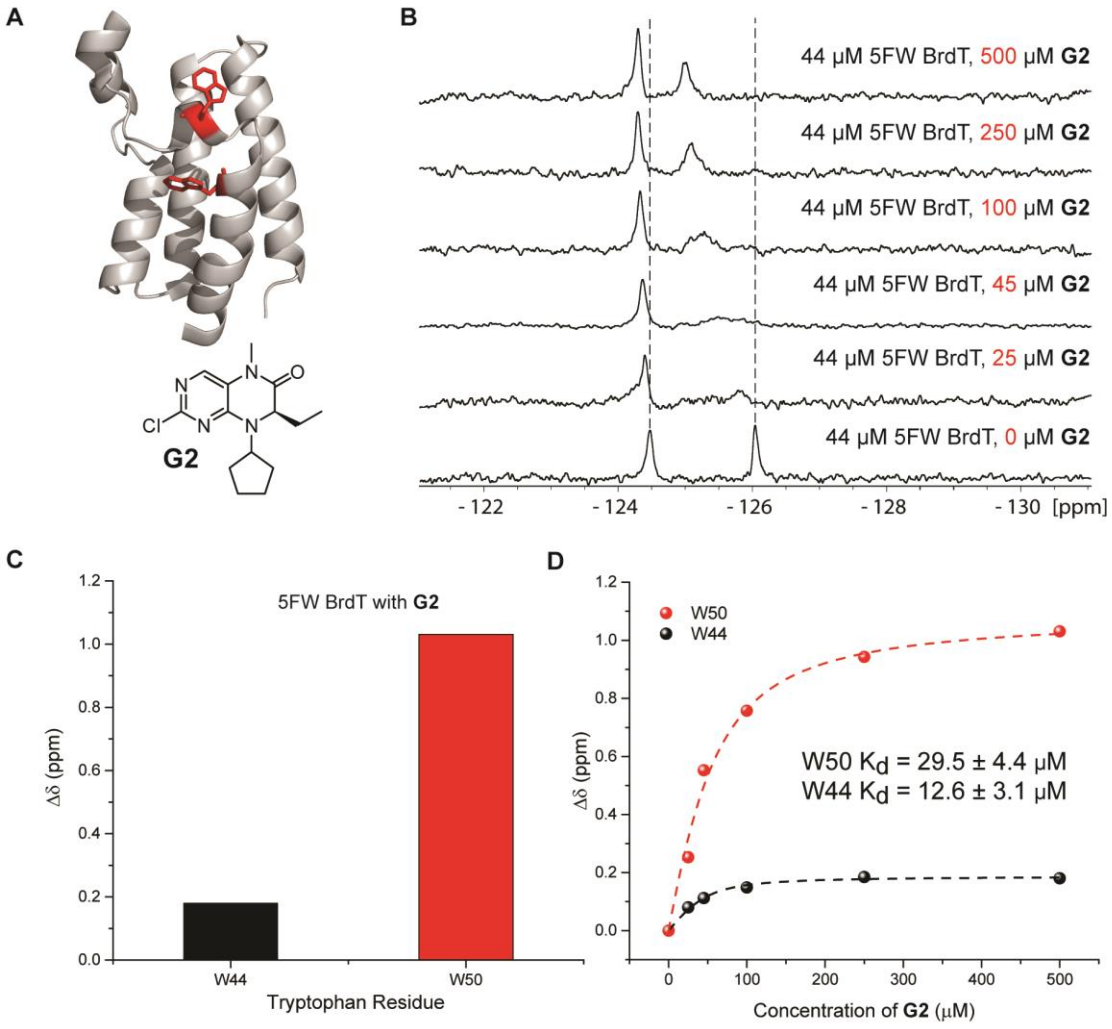


Figure 5.9 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of G2.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **G2**. C) Chemical shift perturbations for all 5FW BrdT resonances at 500 μM **G2**. D) Binding isotherm of W44 and W50 perturbations for the titration of **G2**

Table 5.3 5FW BrdT ^{19}F chemical shift perturbations at varying G2 concentrations

[G2]	Chemical shift (ppm)	W44	W50
0	δ	-124.4795	-126.0442
25	Δδ	0.0804	0.2530
45	Δδ	0.1122	0.5525
100	Δδ	0.1489	0.7572
250	Δδ	0.1849	0.9429
500	Δδ	0.1805	1.0309

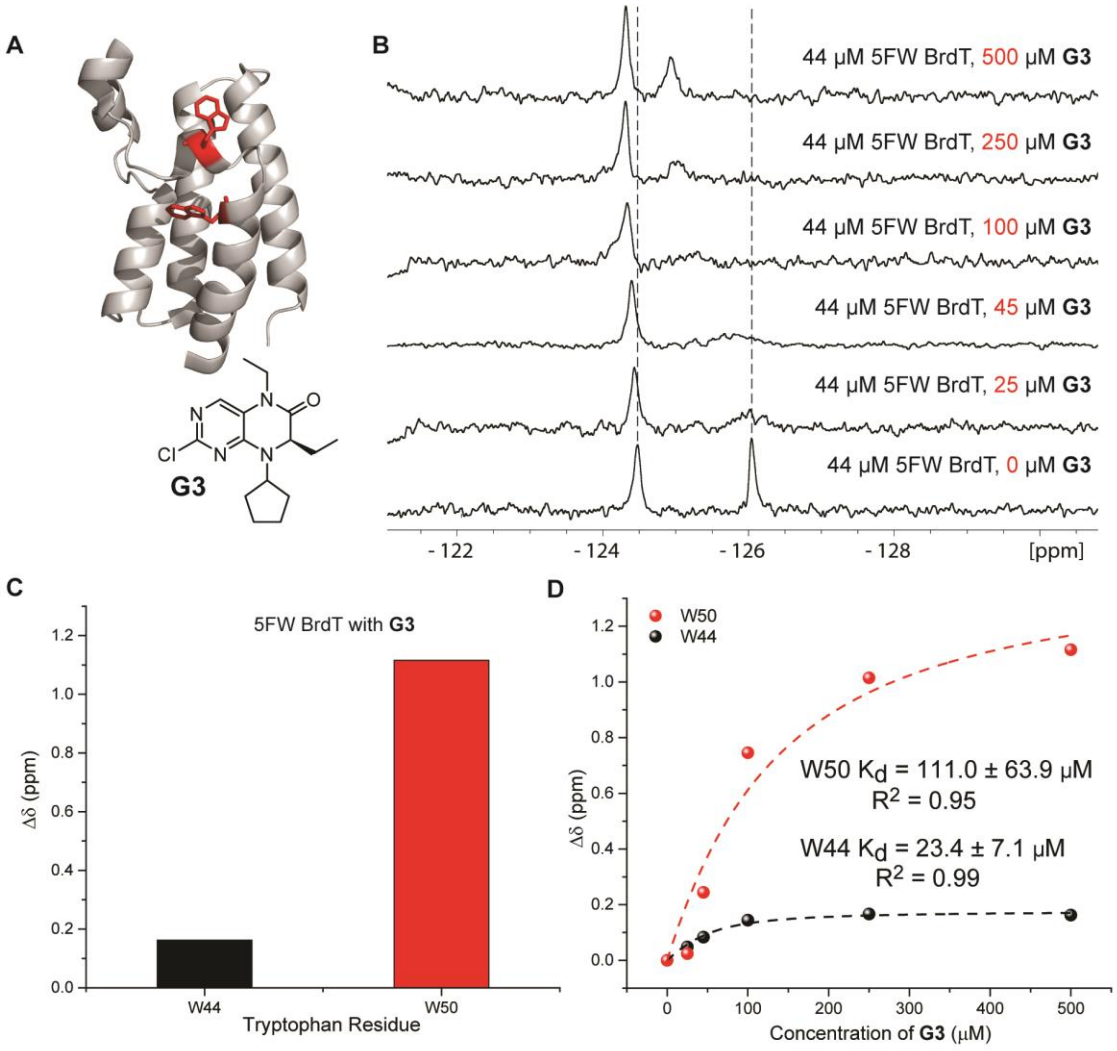


Figure 5.10 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of G3.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **G3**. C) Chemical shift perturbations for all 5FW BrdT resonances at 500 μM **G3**. D) Binding isotherm of W44 and W50 perturbations for the titration of **G3**

Table 5.4 5FW BrdT ^{19}F chemical shift perturbations at varying G3 concentrations

[G3]	Chemical shift (ppm)	W44	W50
0	δ	-124.4795	-126.0442
25	Δδ	0.0478	0.0240
45	Δδ	0.0832	0.2441
100	Δδ	0.1442	0.7457
250	Δδ	0.1663	1.0147
500	Δδ	0.1624	1.1158

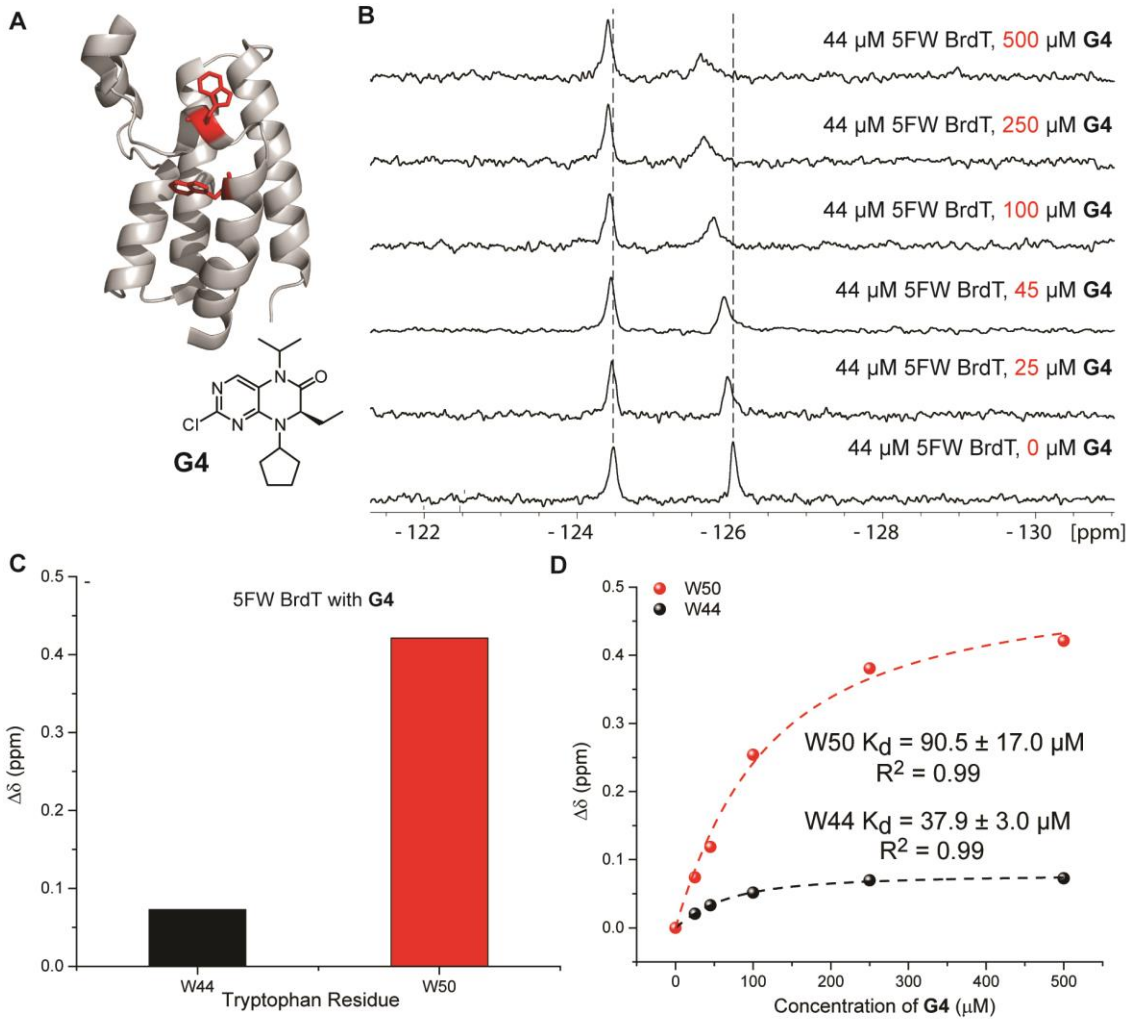


Figure 5.11 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of G4.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **G4**. C) Chemical shift perturbations for all 5FW BrdT resonances at 500 μM **G4**. D) Binding isotherm of W44 and W50 perturbations for the titration of **G4**

Table 5.5 5FW BrdT ^{19}F chemical shift perturbations at varying G4 concentrations

[G4]	Chemical shift (ppm)	W44	W50
0	δ	-124.4795	-126.0442
25	Δδ	0.0207	0.0740
45	Δδ	0.0331	0.1188
100	Δδ	0.0517	0.2541
250	Δδ	0.0697	0.3808
500	Δδ	0.0727	0.4213

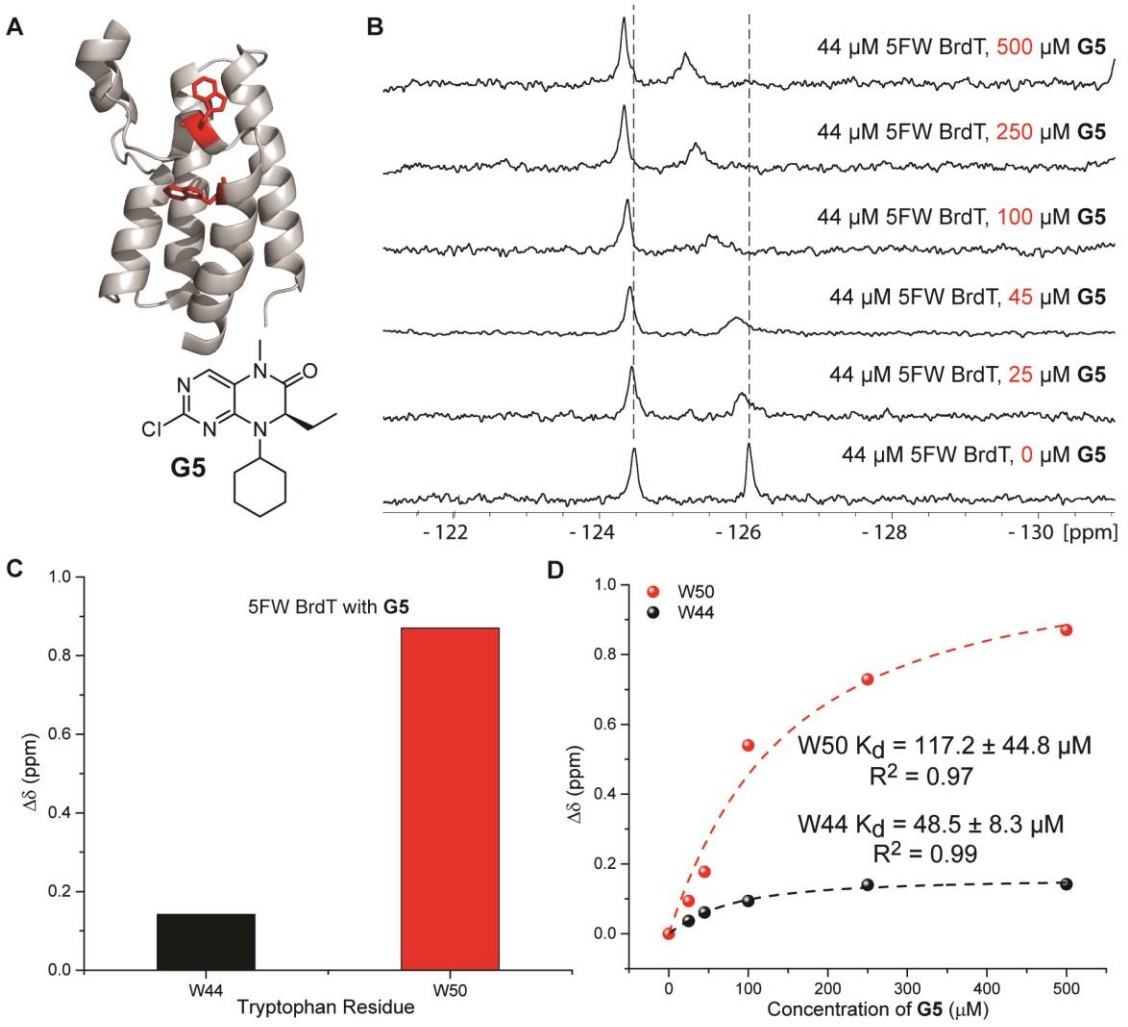


Figure 5.12 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of G5.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **G5**. C) Chemical shift perturbations for all 5FW BrdT resonances at 500 μM **G5**. D) Binding isotherm of W44 and W50 perturbations for the titration of **G5**

Table 5.6 5FW BrdT ^{19}F chemical shift perturbations at varying G5 concentrations

[G5]	Chemical shift (ppm)	W44	W50
0	δ	-124.4795	-126.0442
25	Δδ	0.0367	0.0940
45	Δδ	0.0610	0.1774
100	Δδ	0.0940	0.5399
250	Δδ	0.1399	0.7290
500	Δδ	0.1419	0.8705

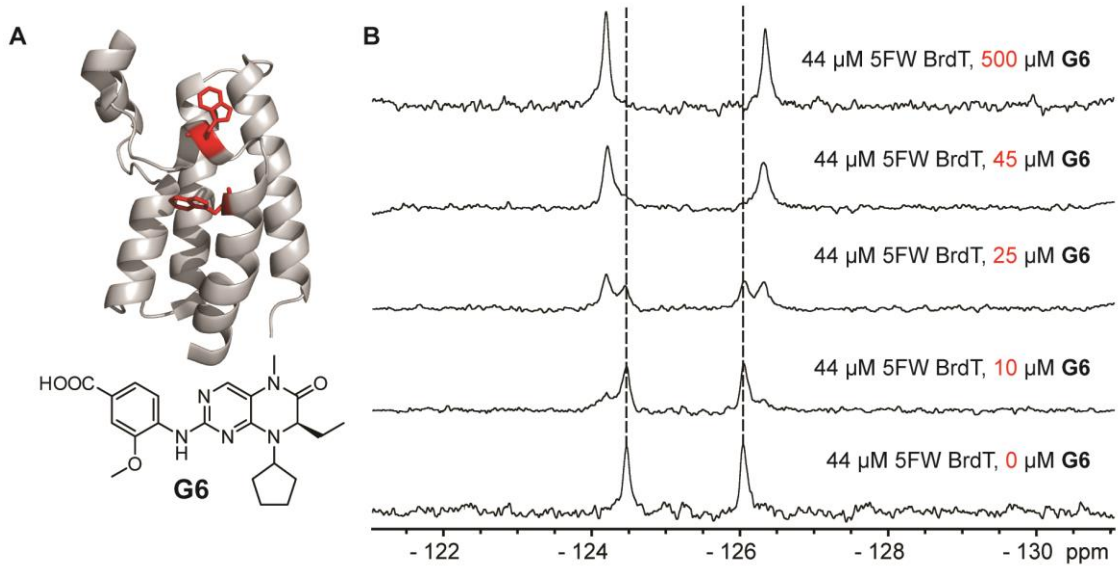
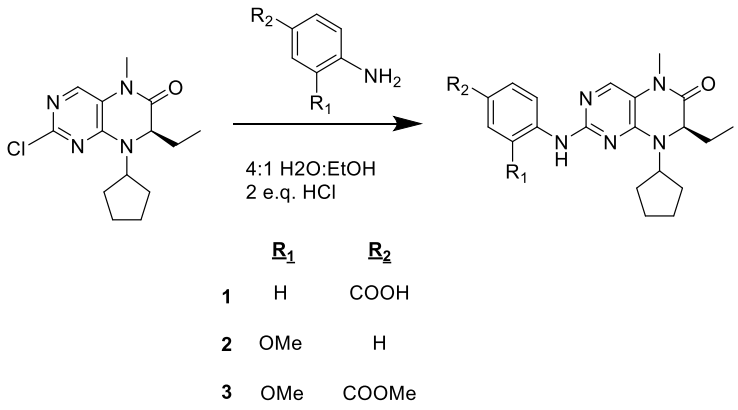


Figure 5.13 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of G6.
A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **G6**.



	R_1	R_2
1	H	COOH
2	OMe	H
3	OMe	COOMe

Scheme 5.1 Synthesis of G6 analogs.

*5.3.3 Synthesis of **G6** analogs*

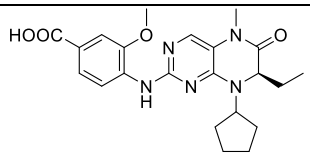
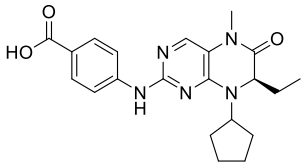
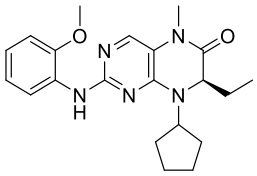
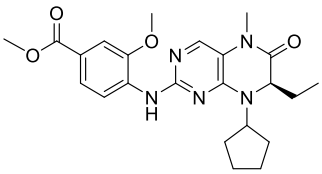
To further explore the SAR of **G6**, I synthesized a new series of analogs (**1**, **2**, and **3**). These three compounds removed the methoxy group on the ring, removed the carboxylic acid completely, and methyl ester-protected the acid, respectively. These compounds were synthesized through a nucleophilic aromatic substitution (S_NAr) reaction between the chloropteridinone and the differently-substituted anilines (Scheme

5.1). Because the affinity of the parent compound **G6** was too tight to be quantified by NMR, their activity was evaluated by other assays. First they were evaluated by DSF. As detailed in Chapter 1, DSF is a thermal shift assay in which the melting temperature (T_m) of a protein is monitored as a function of ligand. Ligands that bind to the protein and stabilize its fold will increase the T_m of the protein. A large positive shift in T_m is indicative of a highly stabilizing ligand. Data is summarized in Table 5.7.This SAR study revealed the importance of the acid functionality on the parent compound **G6**. In removing the methoxy group (**1**), DSF still showed a 6.45 °C stabilization (compared to 8.27 °C for BI-2536). Removing the carboxylic acid group completely (**2**) reduces the change in stabilization to 3.15 °C while methyl esterifying the acid (**3**) lowers it even further to 0.85 °C.

The activity of these ligands was also analyzed by FP and alpha screen with data summarized in Table 5.7. Similar to DSF, FP demonstrated the significance of the carboxylic acid functionality and the lower significance of the methoxy substitution on the ring. By FP, the parent compound **G6** was determined to have a 400 nM IC_{50}. **1** and **3** were measured to have comparable affinities (1.2 and 1.1 μM respectively. By removing the carboxylic acid, **2** exhibited a reduction in affinity by almost an order of magnitude to 5.006 μM. Similarly, Alpha screen data reports almost comparable IC_{50} values for **G6** and **1** while reporting a six-fold worse IC_{50} **2** and only a four-fold worse affinity for **3.** Our studies demonstrated that the parent compound **G6** retained the highest affinity with the removal of either functionality, but especially the carboxylic acid, having negative effects on its affinity for BrdT. Compound **G6** also retains the highest ligand efficiency (LE) and lipophilic ligand efficiency (LLE), which indicate that further optimization should continue from **G6** rather than one of the other analogs

.

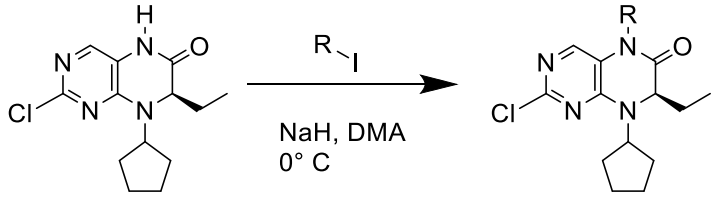
Table 5.7 Summary of data from G6 analogs

Compound	ΔT_m	FP		LE	LLE	α screen IC_{50} (μM)
		IC_{50} (μM)	K_i (FP) (μM)			
G6 	8.27	0.43	< 0.33	0.28	1.7	0.55
1 	6.45	1.2	< 0.33	0.28	1.2	0.64
2 	3.15	5.0	0.76	0.26	0.43	3.5
3 	0.85	1.1	< 0.33	0.25	0.98	2.0
ND - indicates compounds that did not bind, were insoluble, or bound very weakly						

5.3.4 N5-alkyl hypothesis for selectivity

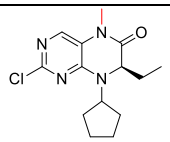
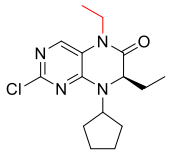
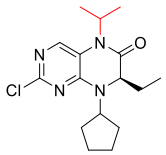
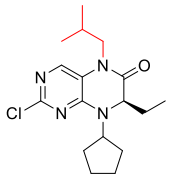
With the goal of developing a selective ligand for BrdT, we developed the hypothesis that one way we may be able to generate selectivity for BrdT over Brd4 using a BI-2536 scaffold would be to take advantage of the N5-alkyl substitution (R_1). We had observed modest reductions in affinity of the 2-chloro-pteridinone analogs containing larger alkyl substitutions at the R_1 position (Table 5.2). However, in comparison to the reductions in affinity that were reported by Chen et al. for their BI-2536 analogs against Brd4, the reductions that we saw were relatively small.[266] Consequently, we tested **G2**, **G3**, and **G4** against 5FW-labeled Brd4 to compare the affinities of the three compounds for both proteins (Figure 5.14, Figure 5.15, Figure 5.16). We also synthesized a fourth compound, **4**, which had an N5-isobutyl substitution (Scheme 5.2). Consistent with what

we expected based on the previous report, the change from methyl to ethyl to isopropyl led to a greater change in magnitude of affinity for Brd4 than it did for BrdT (Table 5.8). While the affinity for BrdT was reduced two- and three-fold for **G3** and **G4**, respectively, the affinity for Brd4 was reduced by nearly an order of magnitude with each additional carbon in **G3** and **G4.** The isobutyl-substituted compound, **4**, had no affinity for either BrdT or Brd4. It is likely that this alkyl substitution is too large and precludes the compound from fitting into the binding pocket.



Scheme 5.2 General alkylation of N5 position on the chloropteridinone scaffold

Table 5.8 Comparison of N5-alkyl analogs for BrdT and Brd4

Compound		BrdT			Brd4		
		K_d	LE	LLE	K_d	LE	LLE
G2		12.6 ± 3.1 29.5 ± 4.4	0.332 0.307	2.28 1.91	< 3	ND	ND
G3		23.4 ± 7.1 111.0 ± 63.9	0.299 0.255	1.67 0.995	2.9 ± 0.9 *	0.357	2.57
G4		37.9 ± 3.0 90.5 ± 17.0	0.272 0.249	1.14 0.763	220 ± 40 55 ± 28	0.225 0.262	-0.212 0.387
4		ND	ND	ND	ND	ND	ND

IC_{50} and K_i values are reported in μM. K_d values from both W44 (top) and W50 (bottom) are noted for BrdT. K_d values for Brd4 were measured using W75 and W81 (bottom) when possible.

* This value is an estimate due to it being over an order of magnitude below the protein concentration.

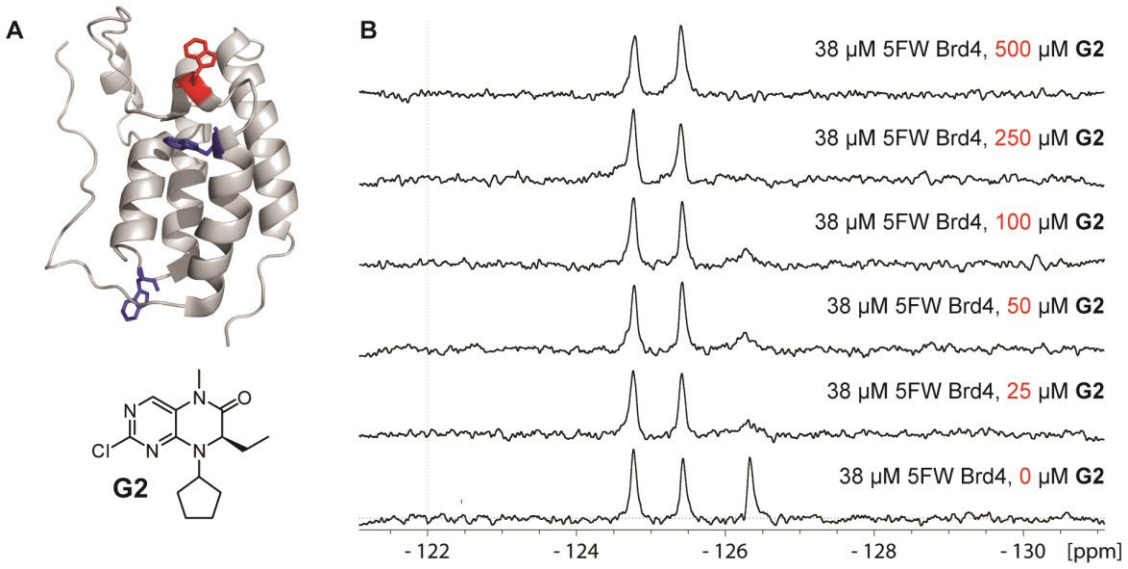


Figure 5.14 ^{19}F NMR spectral analysis of 5FW Brd4 in the presence of increasing concentrations of G2.
A) Ribbon diagram of Brd4 (PDB ID: 3MXF) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW Brd4 with increasing concentrations of **G2**.

Table 5.9 5FW Brd4 ^{19}F chemical shift perturbations at varying G2 concentrations

[G]	Chemical shift (ppm)	W75	W120	W81
0	δ	-124.7678	-125.4327	-126.3300
25	Δδ	0.0054	0.0157	0.0207
45	Δδ	0.0083	0.0107	0.0728
100	Δδ	0.0015	0.0111	0.0544
250	Δδ	0.0011	0.0310	
500	Δδ	0.0189	0.0287	

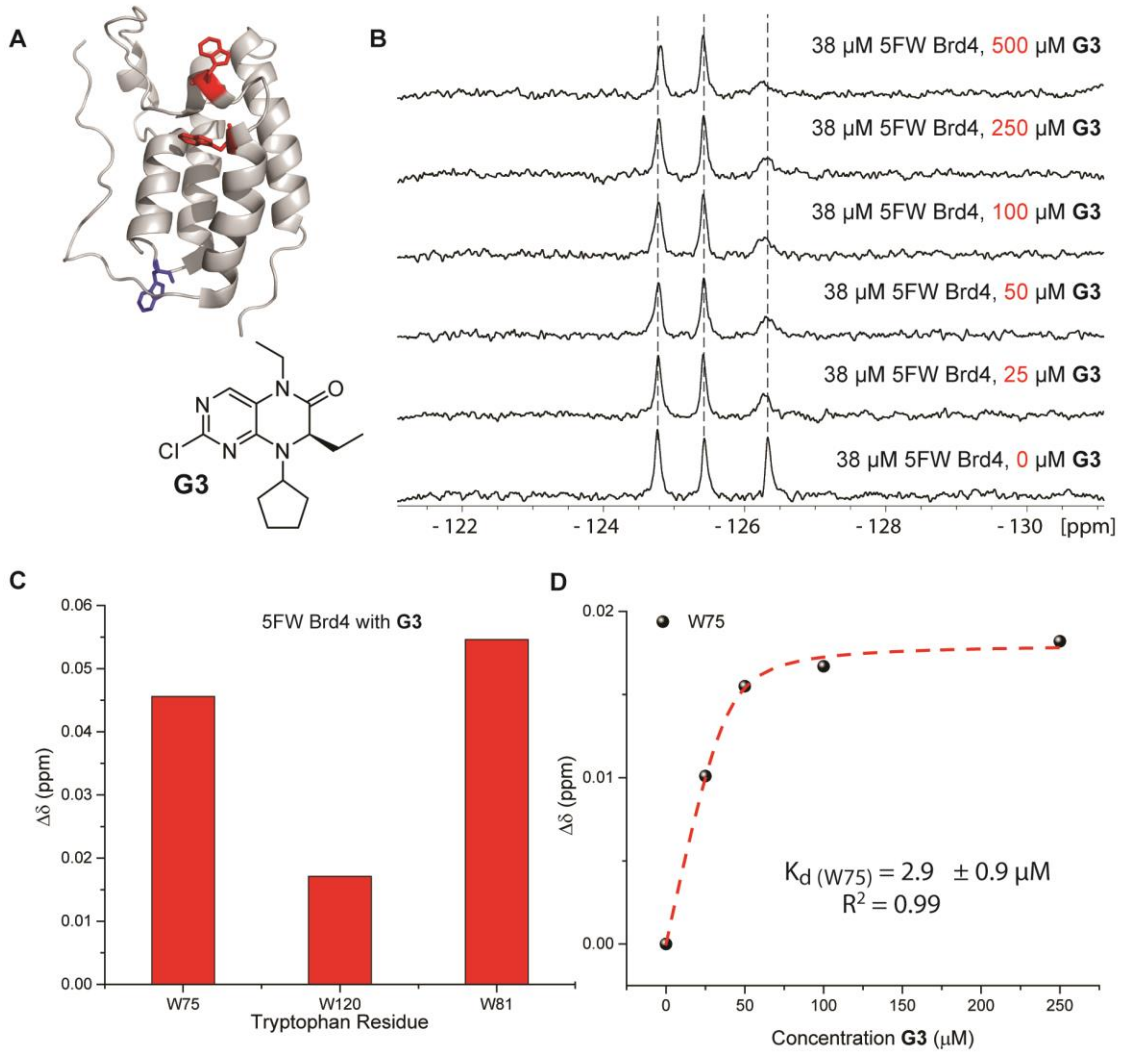


Figure 5.15 ^{19}F NMR spectral analysis of 5FW Brd4 in the presence of increasing concentrations of G3.

A) Ribbon diagram of Brd4 (PDB ID: 3MXF) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW Brd4 with increasing concentrations of **G3**. C) Chemical shift perturbations for all 5FW Brd4 resonances at 500 μM **G3**. D) Binding isotherm of W75 perturbations for the titration of **G3**

Table 5.10 5FW Brd4 ^{19}F chemical shift perturbations at varying G3 concentrations

[G]	Chemical shift (ppm)	W75	W120	W81
0	δ	-124.7678	-125.4327	-126.3300
25	Δδ	0.0101	0.0174	0.0630
45	Δδ	0.0155	0.0095	0.0313
100	Δδ	0.0167	0.0164	0.0449
250	Δδ	0.0182	0.0124	0.0117
500	Δδ	0.0456	0.0171	0.0546

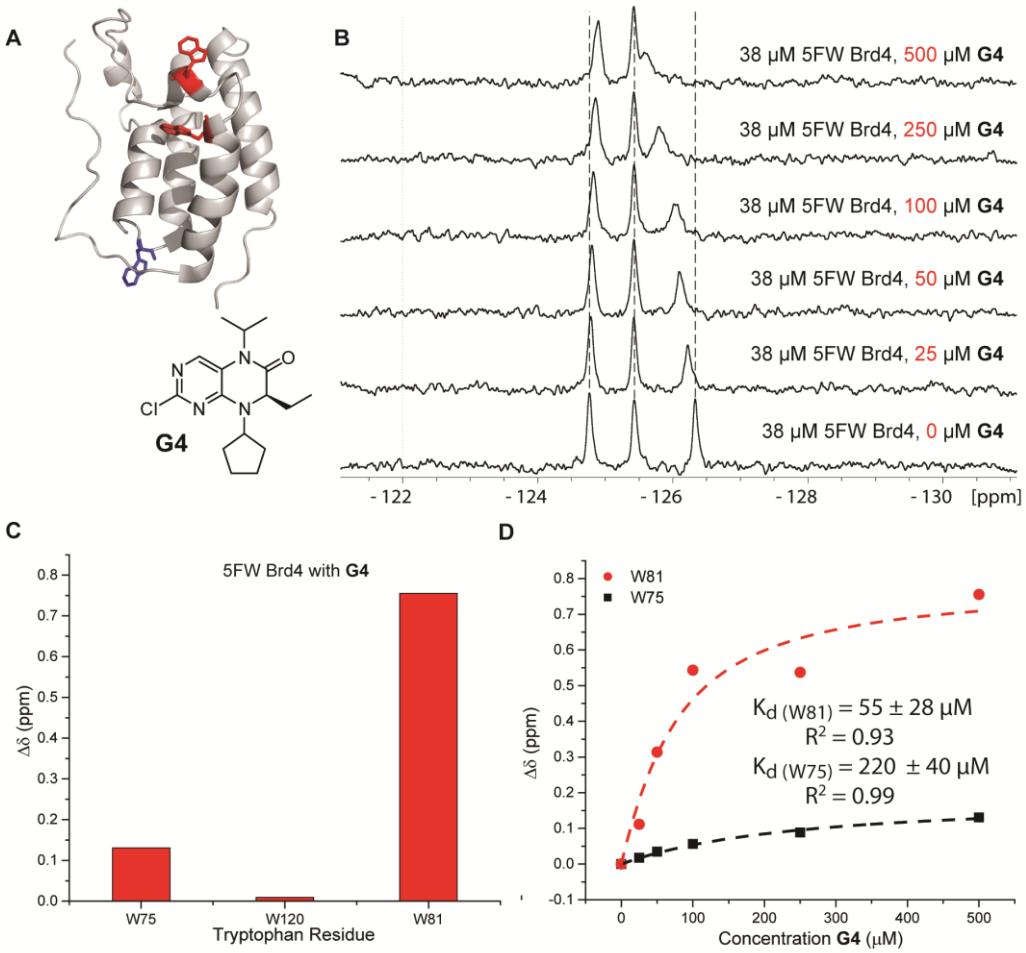


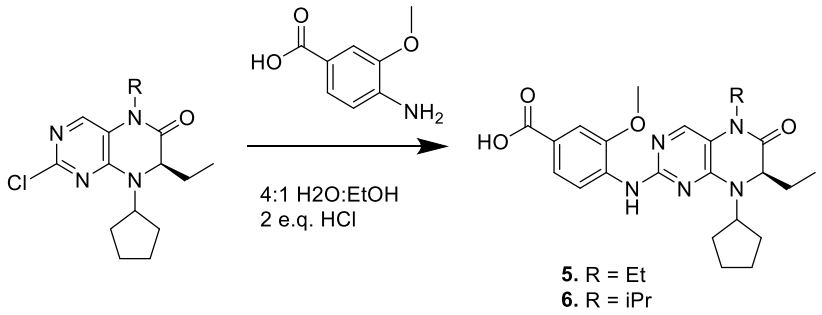
Figure 5.16 ^{19}F NMR spectral analysis of 5FW Brd4 in the presence of increasing concentrations of G4.

A) Ribbon diagram of Brd4 (PDB ID: 3MXF) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW Brd4 with increasing concentrations of **G4**. C) Chemical shift perturbations for all 5FW Brd4 resonances at 500 μM **G4**. D) Binding isotherm of W75 and W81 perturbations for the titration of **G4**

Table 5.11 5FW Brd4 ^{19}F chemical shift perturbations at varying G4 concentrations

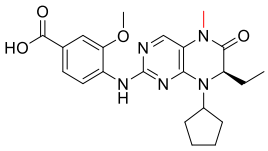
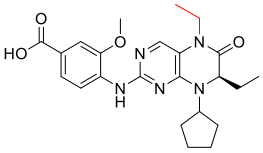
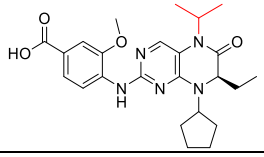
[G]	Chemical shift (ppm)	W75	W120	W81
0	δ	-124.7678	-125.4327	-126.3300
25	Δδ	0.0176	0.0084	0.1113
45	Δδ	0.0347	0.0057	0.3137
100	Δδ	0.0563	0.0040	0.5432
250	Δδ	0.0885	0.0077	0.5373
500	Δδ	0.1306	0.0092	0.7554

To follow this observation, we synthesized two additional **G6** analogs (**5** and **6**) using an S_NAr reaction to further test this N5-alkyl hypothesis for selectivity (Scheme 5.3Table 5.12). Using FP, we compared the affinities of **5** and **6** for both BrdT and Brd4. The methyl-(**G6**) and ethyl-substituted (**5**) still retained a higher affinity for Brd4 than for BrdT (Table 5.12). However, in the change from the ethyl to the isopropyl (**6**), the affinity for Brd4 was reduced by almost an order of magnitude (0.792 μM to 6.624 μM) while the affinity for BrdT was reduced by about a factor of three (1.052 μM to 3.332 μM). These results were encouraging in that it was our first compound with decent potency that had actually changed selectivity to modestly favor BrdT over Brd4.



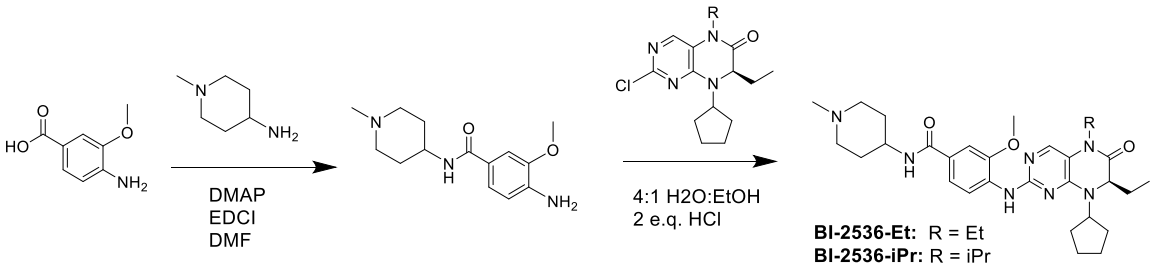
Scheme 5.3 Synthesis of G6 analogs with different N5 alkylations

Table 5.12 Summary of G6 analogs with N5-alkyl substitutions

Compound		BrdT			Brd4		
		FP		α	FP		α
		IC_{50}	K_i	IC_{50}	IC_{50}	K_i	IC_{50}
G6		0.40	< 0.33	0.55	0.77	< 0.33	0.19
5		1.1	< 0.33	0.72	0.79	< 0.33	0.27
6		3.3	0.51	ND	6.6	1.2	ND

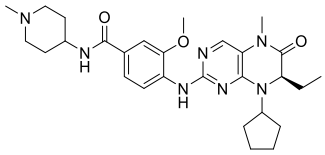
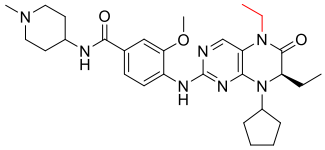
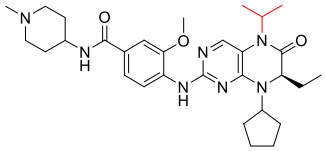
IC_{50} and K_i values are reported in μM

Given these results, we then expanded to the fully elaborated BI-2536 compound with the ethyl and isopropyl substitutions (Scheme 5.4). Surprisingly, FP reported a lower affinity for BrdT than Brd4 for both the ethyl and isopropyl substitution (Table 5.13). These results were unexpected given the change in selectivity that we had seen with the **G6** analogs. This change in affinity could be due to a range of different factors. It has been well documented that in ligand deconstruction studies, smaller fragments or core scaffolds of compounds may bind to the protein target in a different manner than the fully elaborated compound, thereby leading to unexpected results.[267] It is also possible that as the compound is elaborated, the various additions may be inducing small conformational changes in the protein such that the protein is no longer in the same conformation as before.



Scheme 5.4 Synthesis of BI-2536 analogs with different N5 alkylations

Table 5.13 Summary of BI-2536 N5 analogs

Compound	BrdT		Brd4	
	IC_{50}	K_i	IC_{50}	K_i
BI-2536 	0.26*		0.025*	
BI-2536-Et 	0.55	< 0.33	0.41	< 0.33
BI-2536-iPr 	9.4	1.7	7.7	1.4

IC_{50} and K_i values are reported in μM, *Ember et al.[264]

5.3.5 Future Directions for BI-2536 Ligand Deconstruction

While these results were not ideal, they did provide some new insight into the binding properties of BI-2536 for BrdT that had previously not been reported. Further investigation into what may be causing the discrepancy between the G6 analogs and BI-2536 analogs are still needed. From the fragment side, it could be interesting to examine the effects starting from the other end of the compound from the piperidyl ring. It is possible that a fragment from this side of the compound may reveal insight into how that addition affects the overall binding of the elaborated compound. It would also be useful to get crystal structures for all of the 2-chloropteridinone fragments and G6 analogs binding to BrdT and Brd4 to determine if they in fact adopt a different binding conformation than the fully elaborated BI-2536.

5.4 **Fragment Screening for BrdT**

In addition to ligand deconstruction, we took a fragment screening approach to discover and develop BrdT inhibitors. The goal was to discover new chemical scaffolds

that could then be expanded through SAR studies. PrOF NMR was used as a secondary screen to follow up the initial screening hits and to guide the SAR process as new analogs were synthesized and characterized. While many compounds were analyzed in this part of the project, not all of them will be presented here. Rather, the focus of this section is how PrOF NMR was used to help guide the selection and development process. The primary researcher on this screening project was Andrea Wisniewski, and as such, this work in full detail will be published in her dissertation.

5.4.1 DSF for a primary screen

A full fragment screen was performed in the Georg lab using Differential Scanning Fluorimetry (DSF) as the primary screening technique. DSF is a thermal shift assay which reports on changes in protein stability as a function of ligand binding. When a ligand binds to a protein target, the protein is either stabilized or destabilized, thereby increasing or decreasing its melting temperature, T_m. By monitoring the change in T_m by an environmentally sensitive fluorescent dye, ligand binding can be observed. DSF is a rapid technique and requires little material, making it an effective and commonly used screening technique.

Two Rule of 3 (RO3) fragment libraries were obtained from Maybridge, consisting of 2,846 fragment molecules. These compounds were all tested by DSF at a single concentration. Of all 2,846 fragments, 49 met the criteria of a ΔT_m greater than or equal to 0.5 °C. These 49 compounds were then tested in an 8-point dose response experiment. Of these 49 fragments, 24 were confirmed as BrdT hits. The other 25 fragments were removed from further consideration due to poor dose-response or abnormal behavior in the assay.

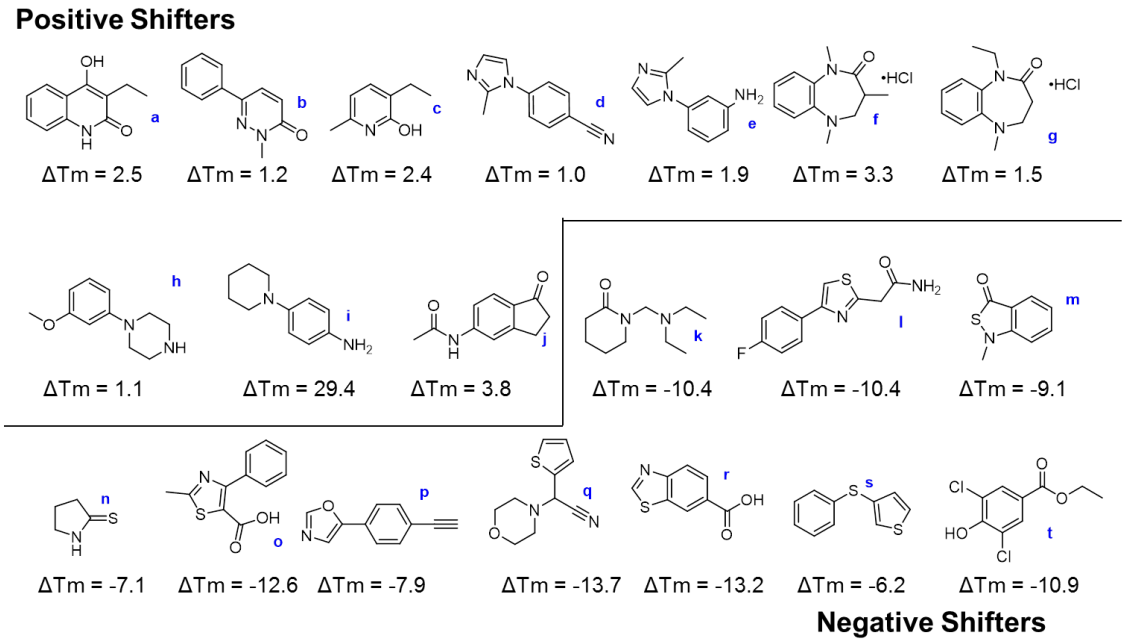


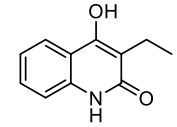
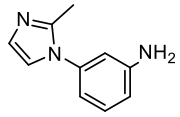
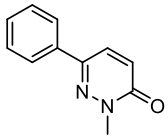
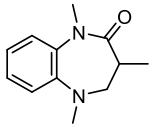
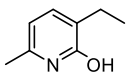
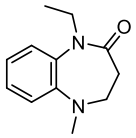
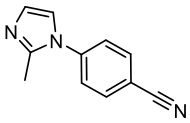
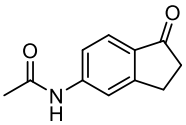
Figure 5.17 Subset of hits from DSF screen that were tested via PrOF NMR.

5.4.2 Validation of hits by PrOF NMR

One of the initial questions immediately following the DSF screen was whether or not stabilizers and destabilizers should both be considered as hits. There is a general consensus that stabilizing compounds are better ligands, but that is not always the case. To address this question, the top ten stabilizing compounds and the top ten destabilizing compounds from the DSF screen were analyzed using PrOF NMR as a validation of binding activity (Figure 5.17). The twenty compounds were tested at 2 mM. Of those compounds, eight of the ten stabilizers and none of the destabilizers yielded a significant change in chemical shift. While it is possible that destabilizing compounds can still be useful, these compounds that were analyzed had demonstrated fairly large negative changes in T_m, which may be indicative more drastic protein unfolding. In answer to the original question, it was decided that compounds with positive ΔT_m values would be favored over compounds with negative ΔT_m values. A negative ΔT_m does not preclude a compound from binding, but specific binding is just one of multiple reasons a compound may generate a negative ΔT_m.

The eight stabilizing compounds that were confirmed as hits by DSF and PrOF NMR were then titrated to determine their K_d, LE, and LLE (Figure 5.18, Figure 5.19, Figure 5.20, Figure 5.21, Figure 5.22, Figure 5.23, Figure 5.24, Figure 5.25), which are all different metrics for prioritizing compounds to explore further (Table 5.14). Given these criteria and the unique indanone scaffold, compound J was selected as the lead compound for optimization. Compounds with LE values greater than 0.3 kcal mol^{-1} HA^{-1} and an LLE value greater than 2 are indicative of good compounds to pursue for further studies.

Table 5.14 Summary of PrOF NMR data from DSF hits

Compound	K_d (μM)	LE	LLE	Compound	K_d (μM)	LE	LLE
A 	83 ± 20	0.40	2.4	**E** 	260 ± 18	0.38	1.8
B 	550 ± 55	0.32	0.60	**F** 	91 ± 14	0.37	2.0
C 	200 ± 38	0.50	2.5	**G** 	31 ± 4.7	0.41	2.4
D 	740 ± 92	0.31	1.8	**J** 	94 ± 7.5	0.39	2.9

K_d values displayed in this table were calculated using either W44 or W50, depending on which yielded a higher R^2 value. K_d values from each both resonances are shown in the titration figures below.

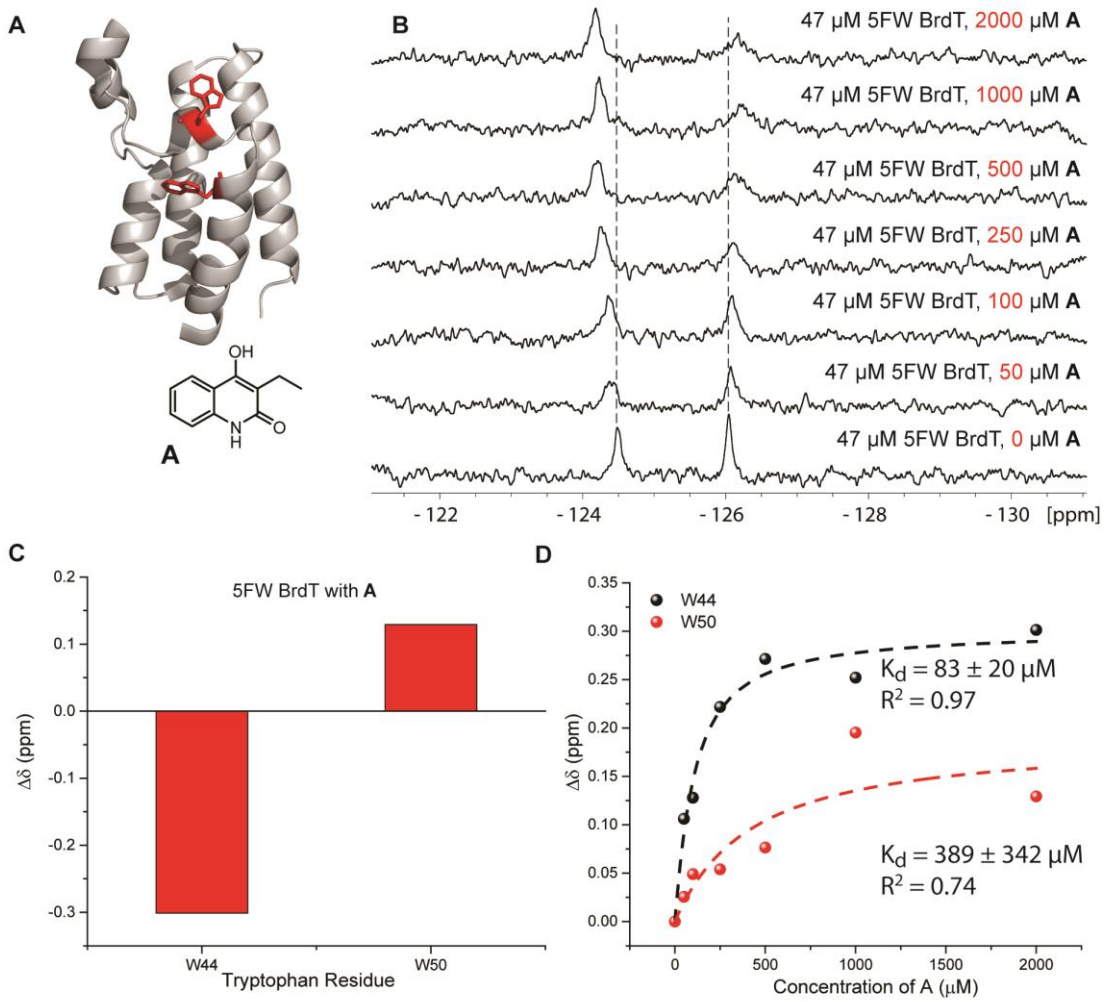


Figure 5.18 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of A.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **A**. C) Chemical shift perturbations for all 5FW BrdT resonances at 2000 μM **A**. D) Binding isotherms for the titration of **A.**

Table 5.15 5FW BrdT ^{19}F chemical shift perturbations at varying A concentrations

[A]	Chemical shift (ppm)	W44	W50
0	δ	-124.4821	-126.0447
50	Δδ	-0.1062	0.0255
100	Δδ	-0.1280	0.0490
250	Δδ	-0.2218	0.0540
500	Δδ	-0.2713	0.0765
1000	Δδ	-0.2520	0.1952
2000	Δδ	-0.3013	0.1293

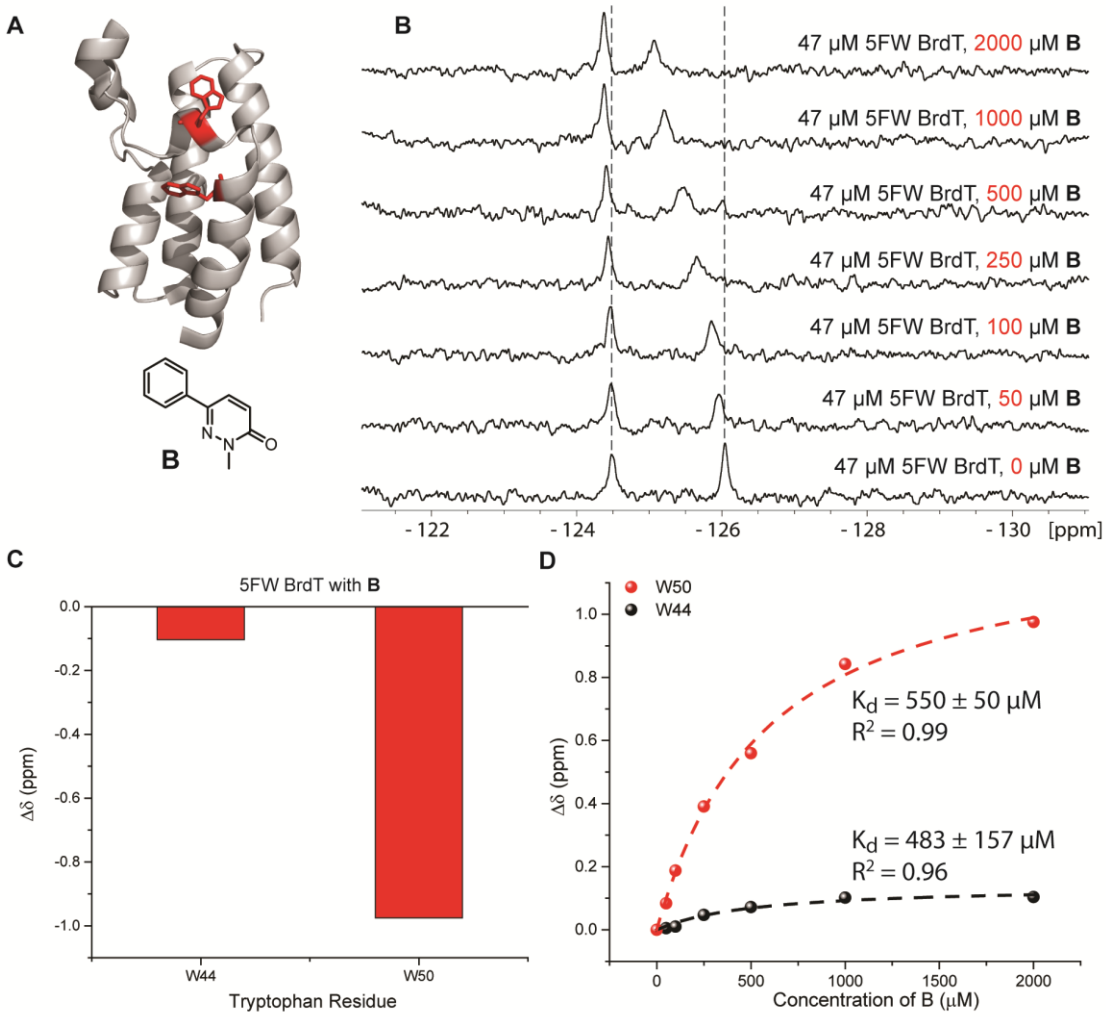


Figure 5.19 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of B.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **B**. C) Chemical shift perturbations for all 5FW BrdT resonances at 2000 μM **B**. D) Binding isotherms for the titration of **B.**

Table 5.16 5FW BrdT ^{19}F chemical shift perturbations at varying B concentrations

[B]	Chemical shift (ppm)	W44	W50
0	δ	-124.4821	-126.0447
50	Δδ	-0.0049	-0.0837
100	Δδ	-0.0099	-0.1876
250	Δδ	-0.0470	-0.3907
500	Δδ	-0.0719	-0.5597
1000	Δδ	-0.1019	-0.8426
2000	Δδ	-0.1037	-0.9756

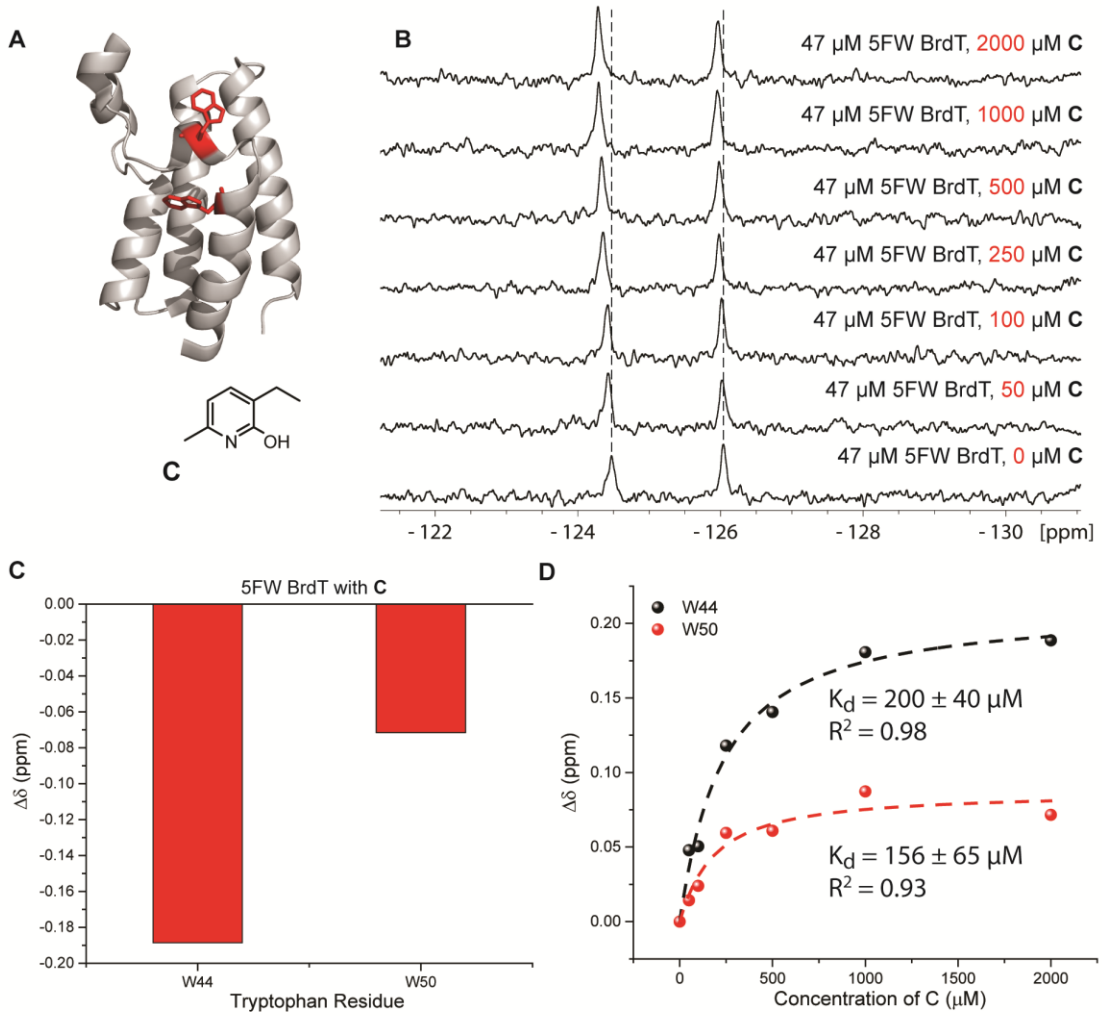


Figure 5.20 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of C.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **C**. C) Chemical shift perturbations for all 5FW BrdT resonances at 2000 μM **C** D) Binding isotherms for the titration of **C.**

Table 5.17 5FW BrdT ^{19}F chemical shift perturbations at varying C concentrations

[C]	Chemical shift (ppm)	W44	W50
0	δ	-124.4729	-126.0409
50	Δδ	-0.0479	-0.0143
100	Δδ	-0.0505	-0.0240
250	Δδ	-0.1180	-0.0595
500	Δδ	-0.1406	-0.0609
1000	Δδ	-0.1807	-0.0873
2000	Δδ	-0.1886	-0.0716

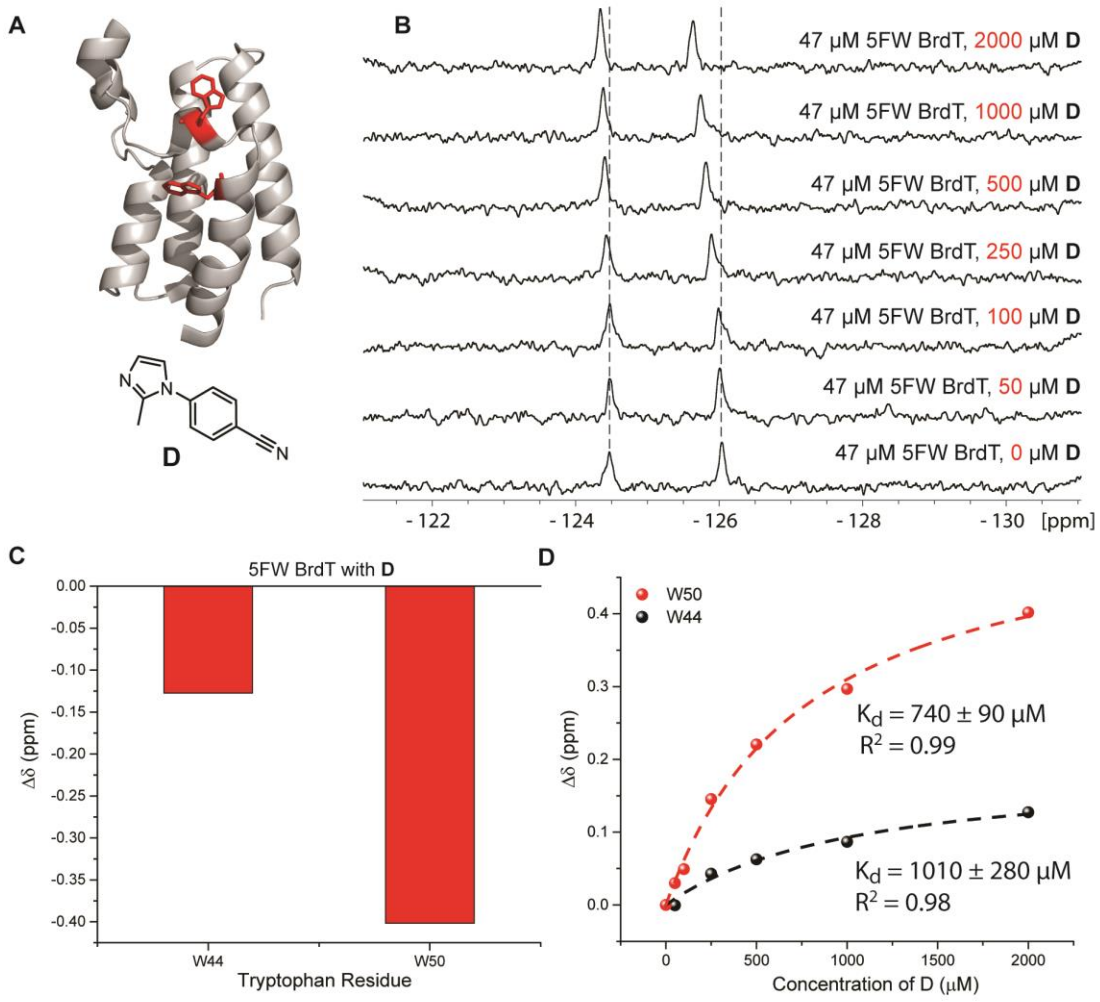


Figure 5.21 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of D.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **D**. C) Chemical shift perturbations for all 5FW BrdT resonances at 2000 μM **D**. D) Binding isotherms for the titration of **D.**

Table 5.18 5FW BrdT ^{19}F chemical shift perturbations at varying D concentrations

[D]	Chemical shift (ppm)	W44	W50
0	δ	-124.4729	-126.0409
50	Δδ	0.0005	-0.0299
100	Δδ	0.0045	-0.0492
250	Δδ	-0.0429	-0.1454
500	Δδ	-0.0627	-0.2205
1000	Δδ	-0.0867	-0.2967
2000	Δδ	-0.1273	-0.4017

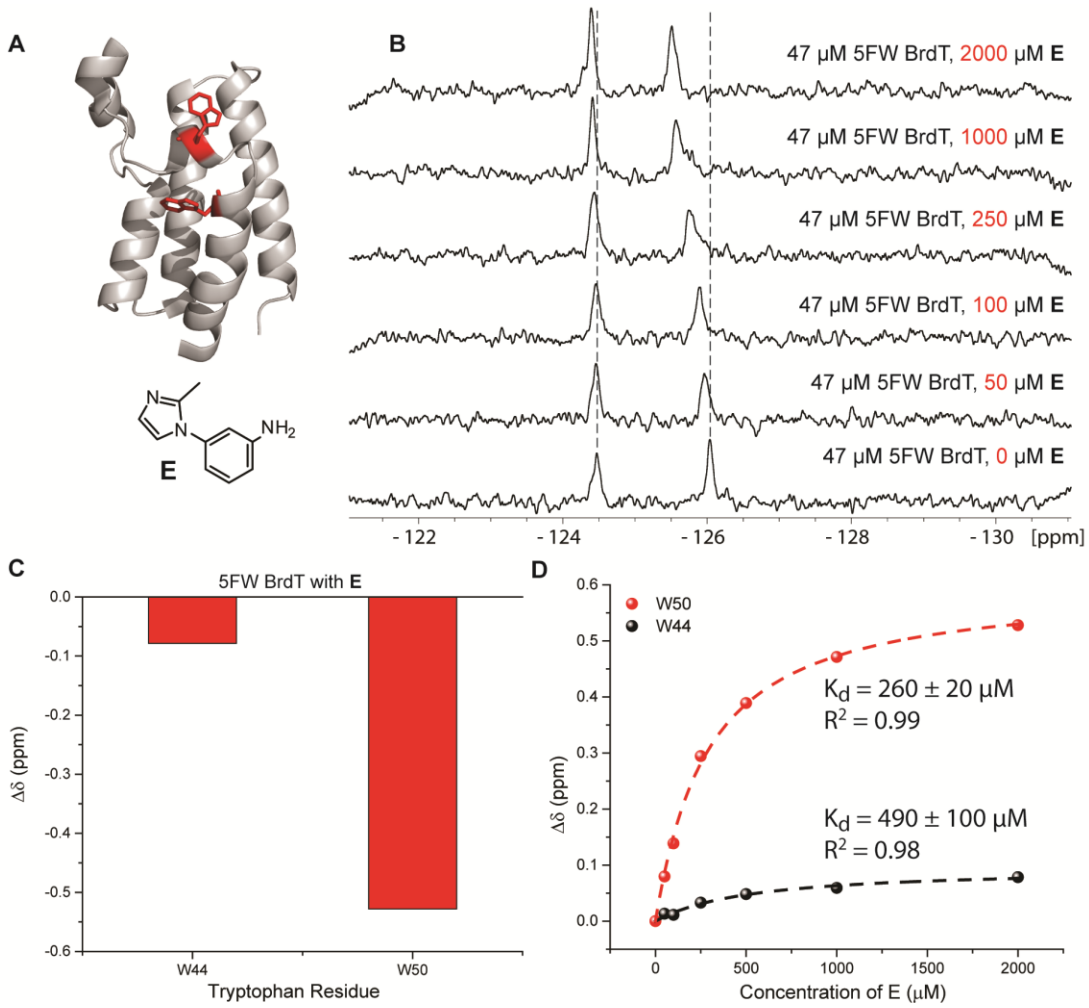


Figure 5.22 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of E.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **E**. C) Chemical shift perturbations for all 5FW BrdT resonances at 2000 μM **E**. D) Binding isotherms for the titration of **E.**

Table 5.19 5FW BrdT ^{19}F chemical shift perturbations at varying E concentrations

[E]	Chemical shift (ppm)	W44	W50
0	δ	-124.4729	-126.0409
50	Δδ	-0.0133	-0.0797
100	Δδ	-0.0112	-0.1388
250	Δδ	-0.0330	-0.2944
500	Δδ	-0.0483	-0.3890
1000	Δδ	-0.0592	-0.4713
2000	Δδ	-0.0785	-0.5278

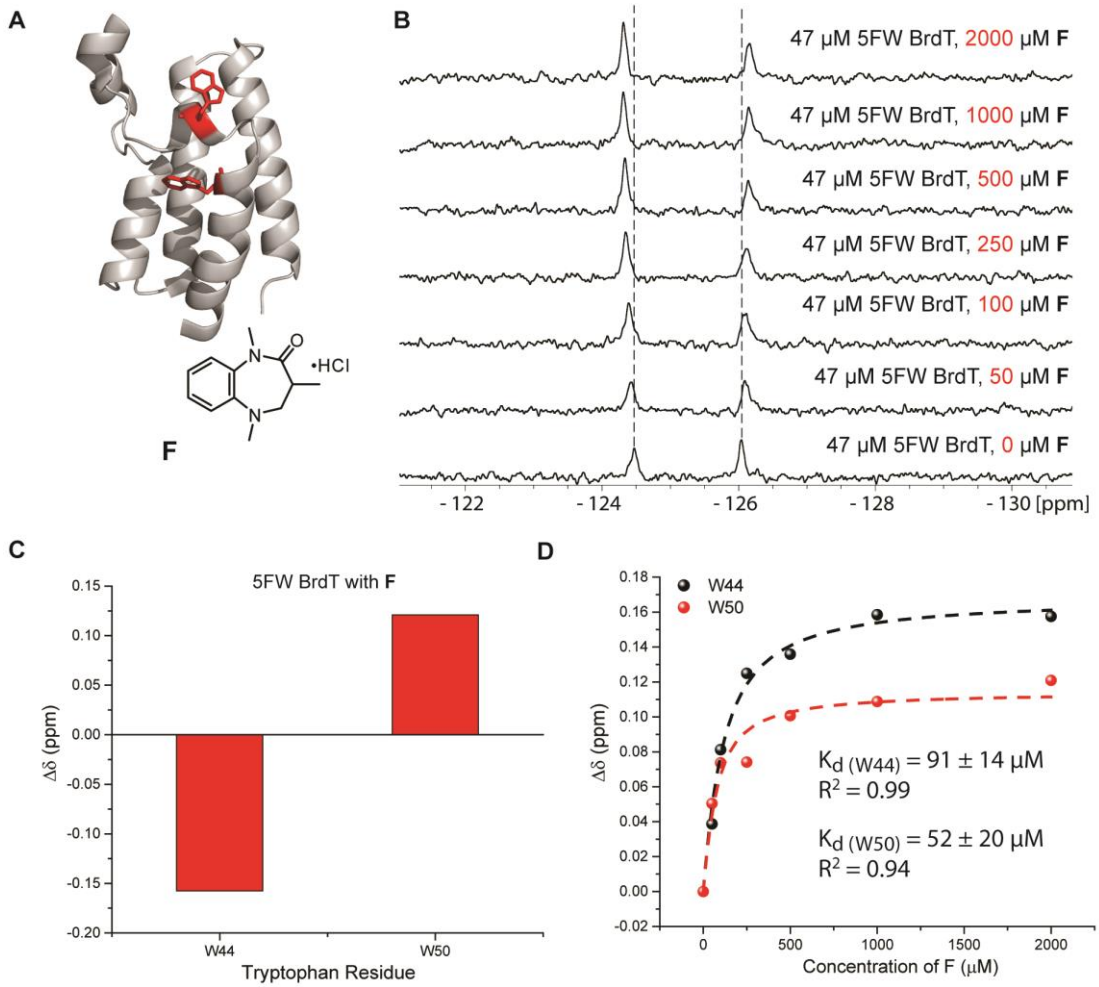


Figure 5.23 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of F.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **F**. C) Chemical shift perturbations for all 5FW BrdT resonances at 2000 μM **F**. D) Binding isotherms for the titration of **F.**

Table 5.20 5FW BrdT ^{19}F chemical shift perturbations at varying F concentrations

[F]	Chemical shift (ppm)	W44	W50
0	δ	-124.4729	-126.0409
50	Δδ	-0.0387	0.0504
100	Δδ	-0.0812	0.0737
250	Δδ	-0.1249	0.0741
500	Δδ	-0.1359	0.1007
1000	Δδ	-0.1585	0.1088
2000	Δδ	-0.1575	0.1209

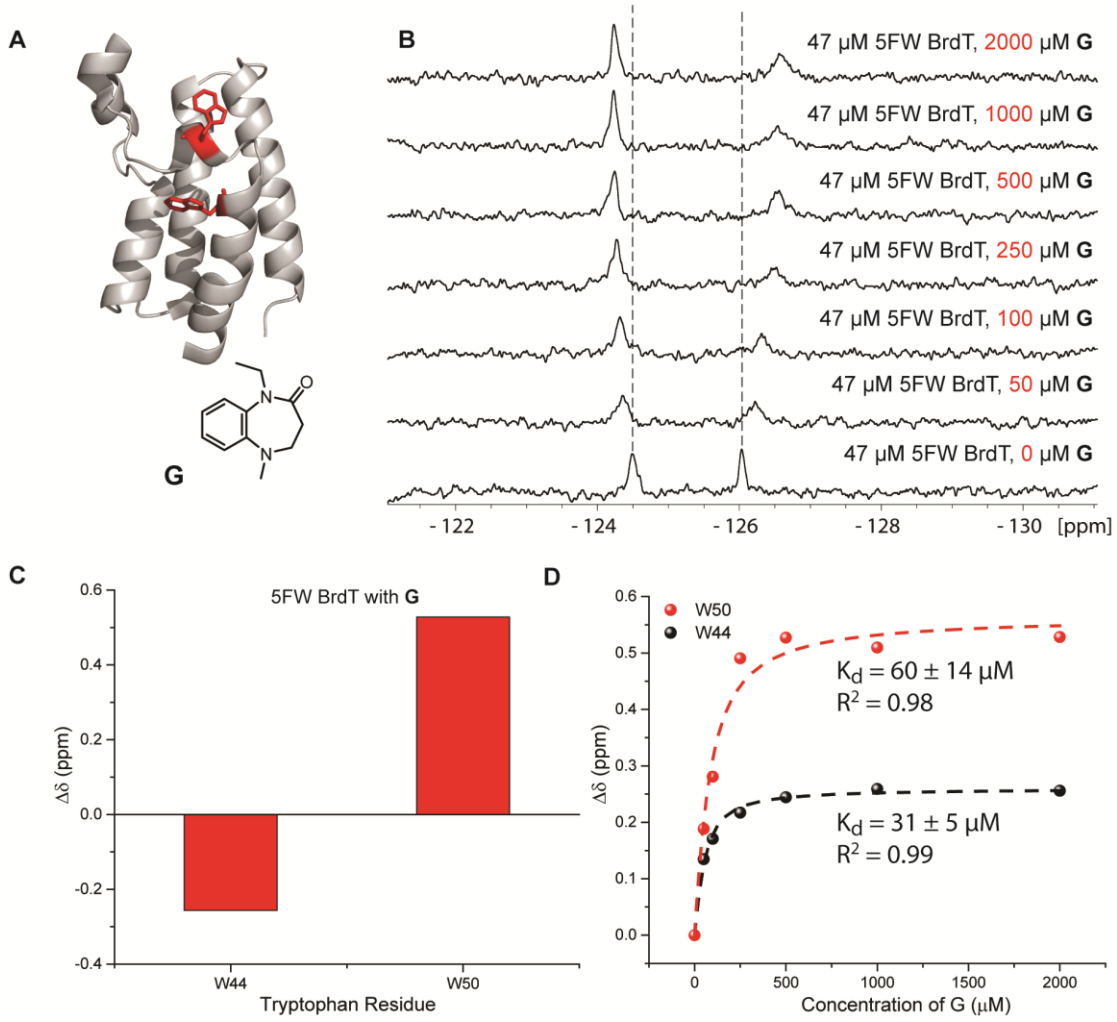


Figure 5.24 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of G.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **G**. C) Chemical shift perturbations for all 5FW BrdT resonances at 2000 μM **G**. D) Binding isotherms for the titration of **G.**

Table 5.21 5FW BrdT ^{19}F chemical shift perturbations at varying G concentrations

[G]	Chemical shift (ppm)	W44	W50
0	δ	-124.4888	-126.0330
50	Δδ	-0.1348	0.1885
100	Δδ	-0.1708	0.2808
250	Δδ	-0.2172	0.4905
500	Δδ	-0.2443	0.5272
1000	Δδ	-0.2591	0.5097
2000	Δδ	-0.2560	0.5282

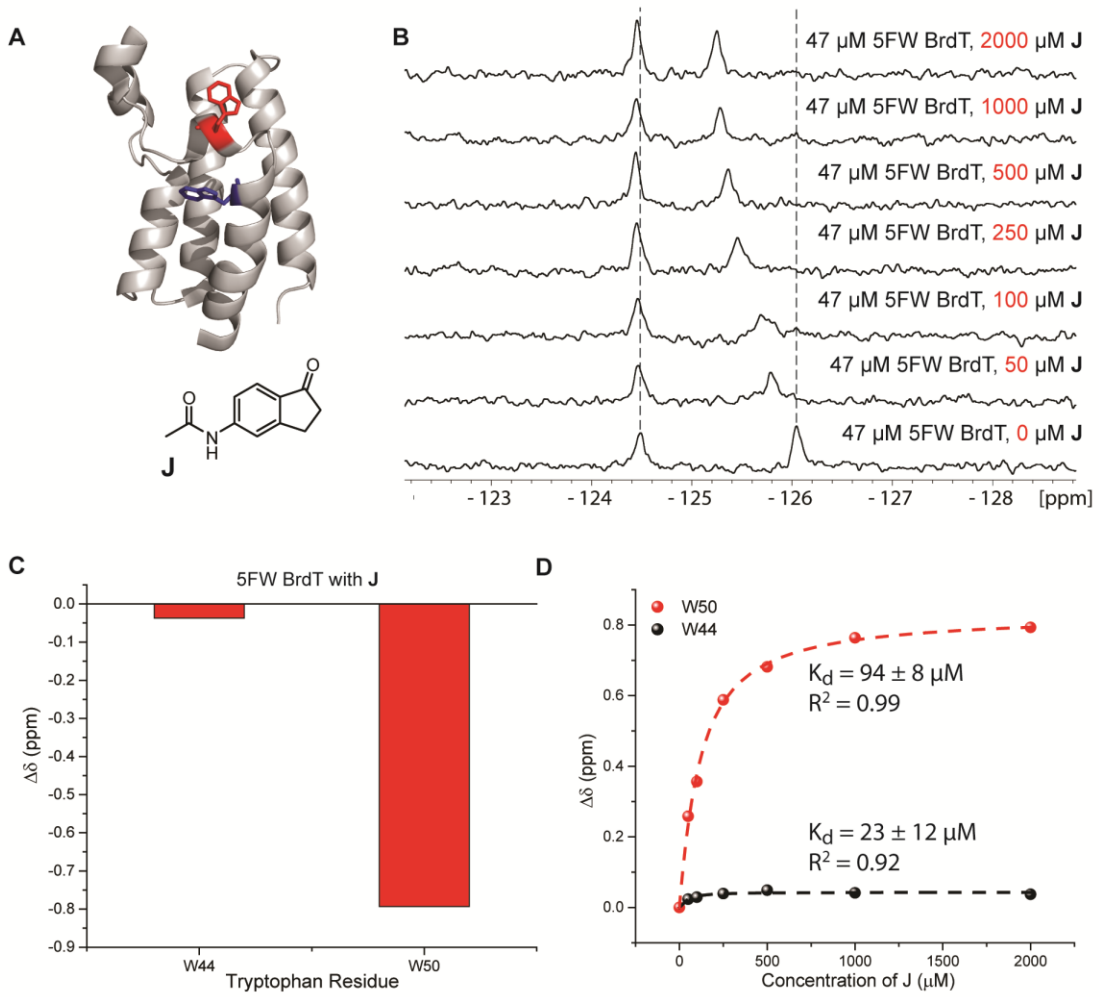


Figure 5.25 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of J.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **J**. C) Chemical shift perturbations for all 5FW BrdT resonances at 2000 μM **J**. D) Binding for the titration of **J.**

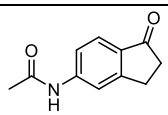
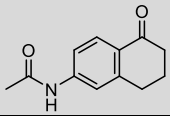
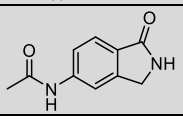
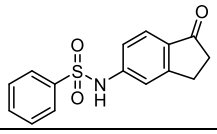
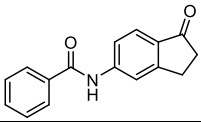
Table 5.22 5FW BrdT ^{19}F chemical shift perturbations at varying J concentrations

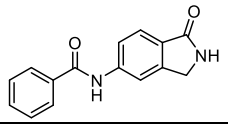
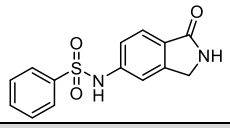
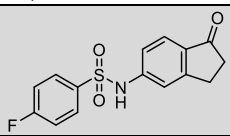
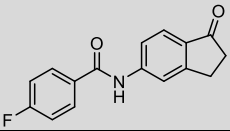
[J]	Chemical shift (ppm)	W44	W50
0	δ	-124.4888	-126.0330
50	Δδ	-0.0214	-0.2457
100	Δδ	-0.0272	-0.3440
250	Δδ	-0.0372	-0.5755
500	Δδ	-0.0469	-0.6690
1000	Δδ	-0.0391	-0.7513
2000	Δδ	-0.0352	-0.7808

5.4.3 SAR development of lead compound

The lead compound, **J**, was derivatized, primarily near the amide, though some changes were also made to the indanone five-membered ring as well. Eight of the J analogs (**J1-J8**) are shown below in Table 5.23 and in the titration figures (Figure 5.26, Figure 5.27, Figure 5.28, Figure 5.29, Figure 5.30, Figure 5.31, Figure 5.32) Titration results showed the indanone in **J** to be important as a naphthalenone (**J1**) and an isoindolinone (**J2**) demonstrated poor binding (Table 5.23). The largest improvement in affinity came from a replacement of the amide with a sulfonamide as noted with **J3** and **J7** with a K_d values of 4.9 and 3.6 μM, respectively. Closely related compounds that lack the sulfonamide (**J4**, **J5**, and **J8**) or contain the unfavorable isoindolinone core (**J5** and **J6**) have weaker affinities and lower LE and LLE values. Similar results were obtained using both FP and an alpha-screen competition assay where **J3** and **J7** both have single digit micromolar IC_{50} values. Additional synthesis to further develop these ligands is ongoing.

Table 5.23 Summary of compound J analogs

Compound		DSF	PrOF NMR			FP	α-screen
		ΔT_m (°C)	**K_d (μM)**	**LE**	**LLE**	**IC_{50} (μM)**	**IC_{50} (μM)**
J		3.8	95 ± 8	0.39	2.31	240	> 50
J1		ND	460 ± 50	0.30	1.9	ND	ND
J2		0.4	ND	ND	ND	NB	> 50
J3		4.28	4.9 ± 0.5	0.36	3.76	3.3	6.28
J4		1.96	50 ± 7	0.31	1.80	190	44

J5		1.66	117 ± 28	0.28	2.13	41	> 50
J6		2.28	24 ± 0.3	0.31	3.66	19	33
J7		3.36	3.6 ± 0.6	0.33	3.56	8.2	6.1
J8		ND	52 ± 13	0.29	1.56	86	> 50
ND denotes measurements that were not determined due to assay conditions, solubility, or lack of activity NB denotes samples that exhibited no binding behavior							

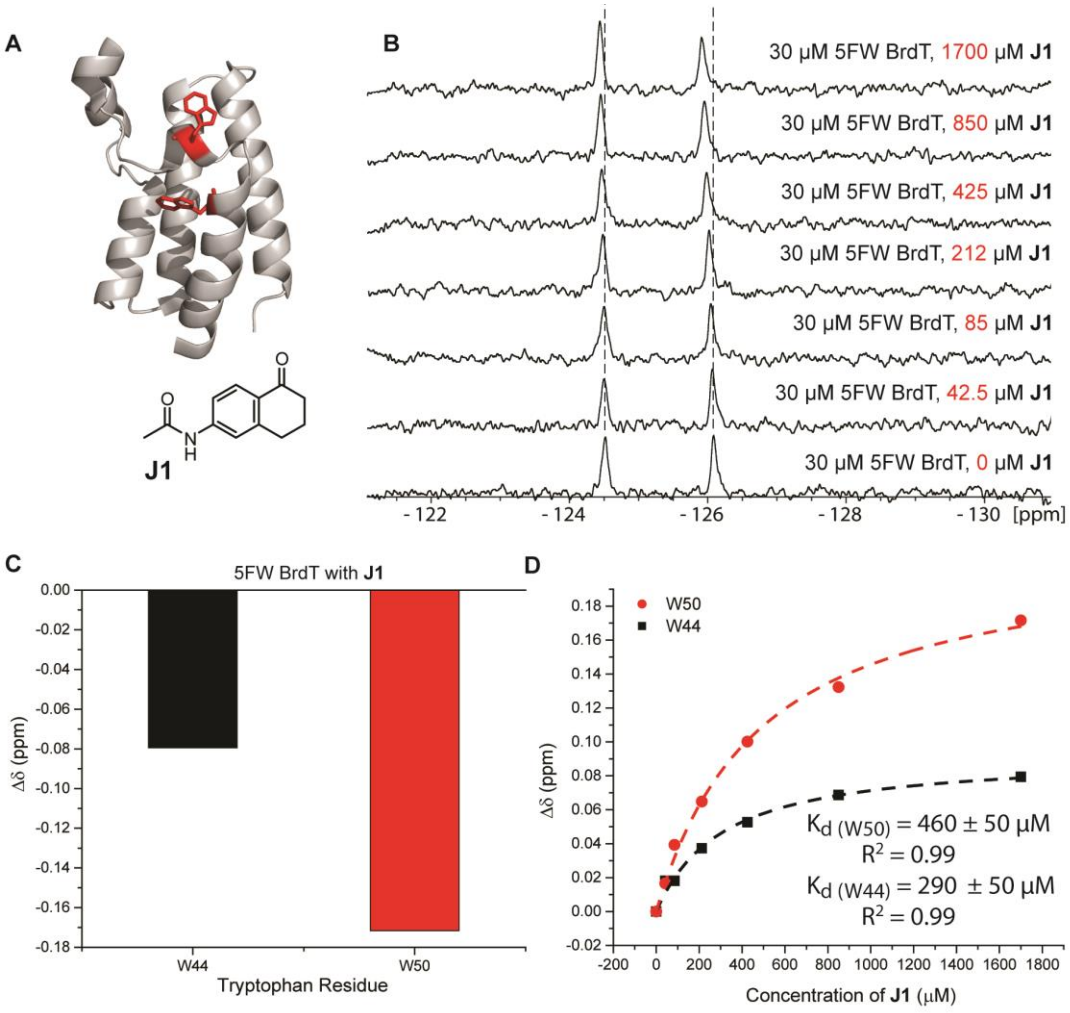


Figure 5.26 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of J1.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **J1**. C) Chemical shift perturbations for all 5FW BrdT resonances at 1700 μM **J1**. D) Binding isotherm of W50 and W44 perturbations for the titration of **J1**

Table 5.24 5FW BrdT ^{19}F chemical shift perturbations at varying J1 concentrations

[J1]	Chemical shift (ppm)	W44	W50
0	δ	-124.5218	-126.0854
42.5	Δδ	-0.0181	-0.0166
85	Δδ	-0.0181	-0.0393
212	Δδ	-0.0373	-0.0648
425	Δδ	-0.0526	-0.1001
850	Δδ	-0.0687	-0.1323
1700	Δδ	-0.0794	-0.1716

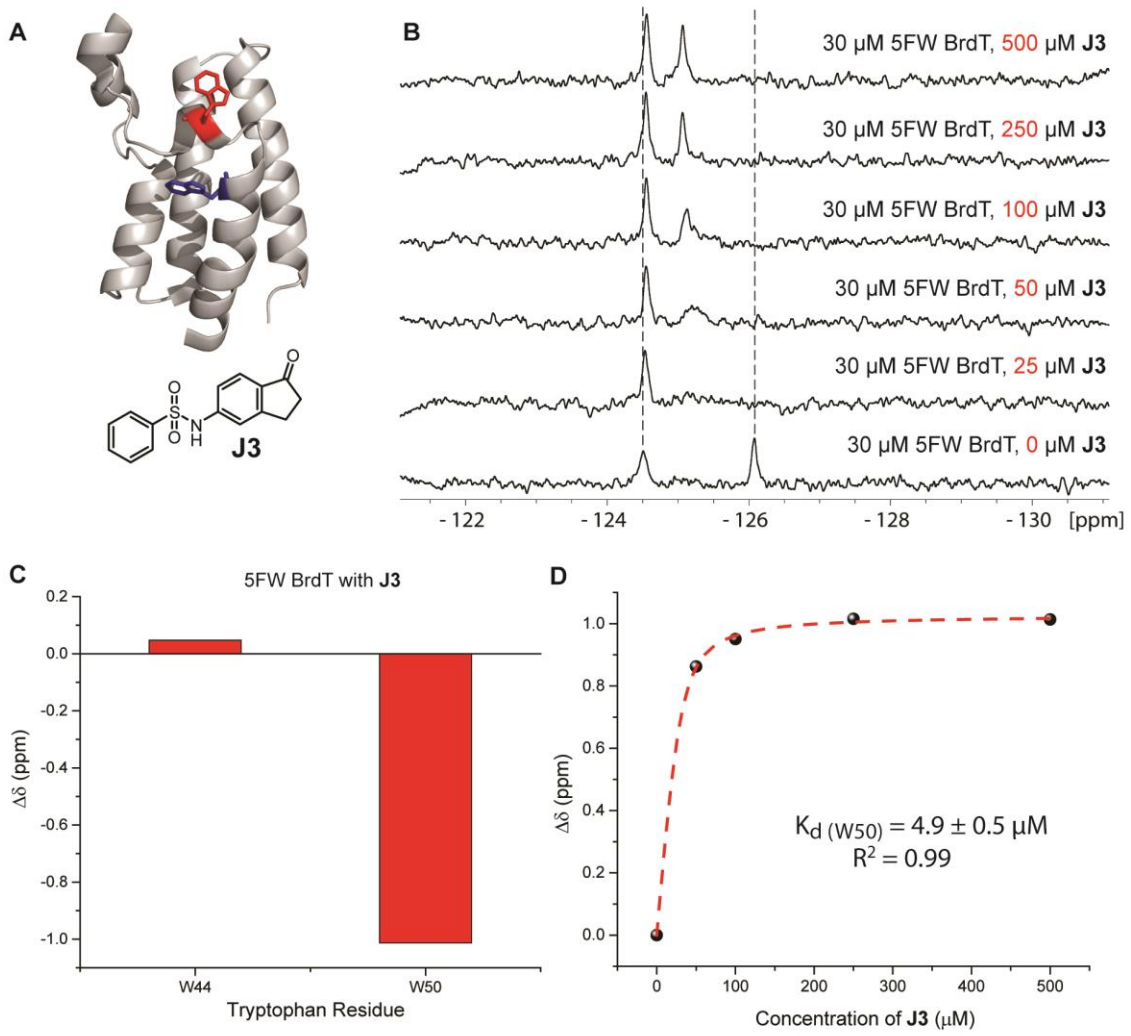


Figure 5.27 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of J3.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **J3**. C) Chemical shift perturbations for all 5FW BrdT resonances at 500 μM **J3**. D) Binding isotherm of W50 perturbations for the titration of **J3**

Table 5.25 5FW BrdT ^{19}F chemical shift perturbations at varying J3 concentrations

[J3]	Chemical shift (ppm)	W44	W50
0	δ	-124.5098	-126.0815
25	Δδ	0.0257	
50	Δδ	0.0398	-0.8625
100	Δδ	0.0440	-0.9508
250	Δδ	0.0395	-1.0157
500	Δδ	0.0477	-1.0137

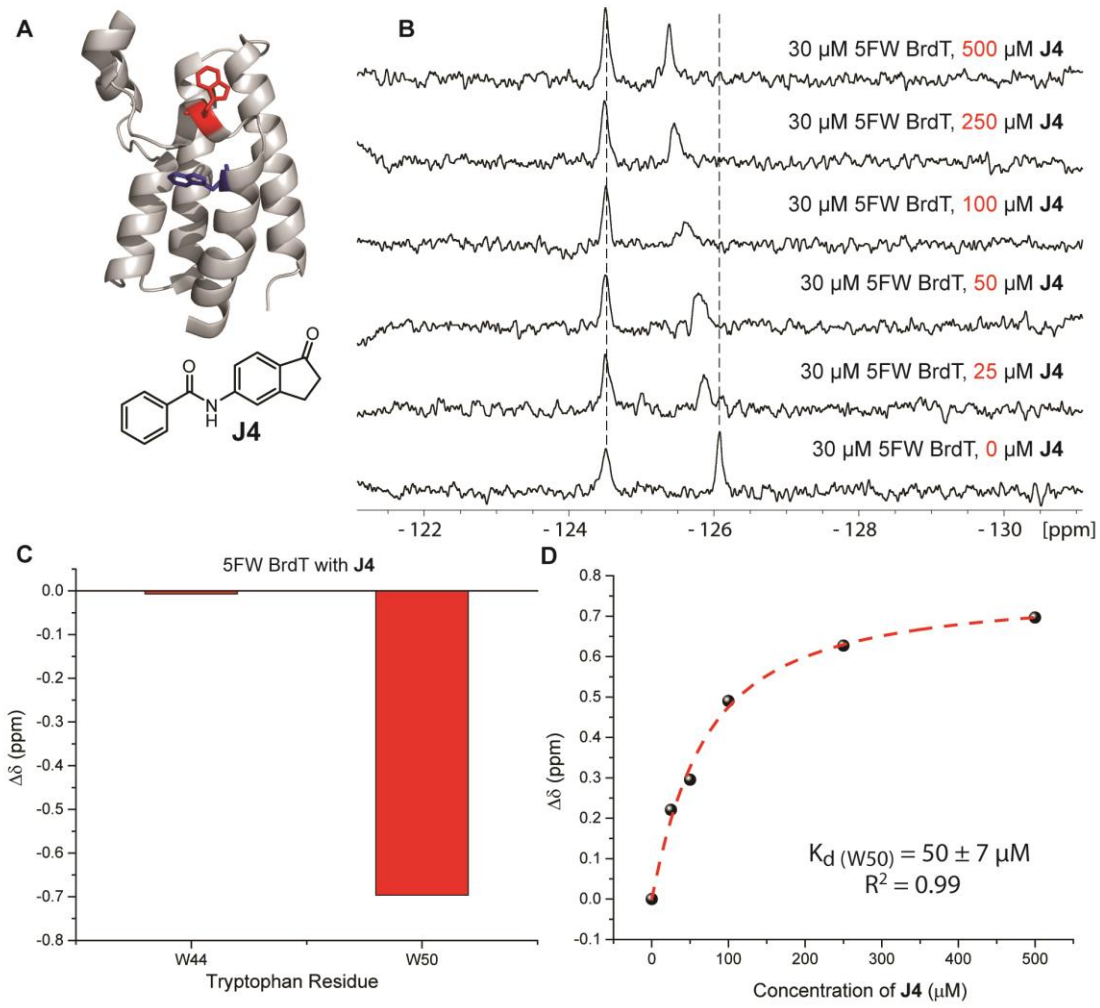


Figure 5.28 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of J4.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **J4**. C) Chemical shift perturbations for all 5FW BrdT resonances at 500 μM **J4**. D) Binding isotherm of W50 perturbations for the titration of **J4**

Table 5.26 5FW BrdT ^{19}F chemical shift perturbations at varying J4 concentrations

[J4]	Chemical shift (ppm)	W44	W50
0	δ	-124.5098	-126.0815
25	Δδ	-0.0105	-0.2208
50	Δδ	-0.0072	-0.2953
100	Δδ	0.0017	-0.4900
250	Δδ	-0.0189	-0.6271
500	Δδ	-0.0069	-0.6965

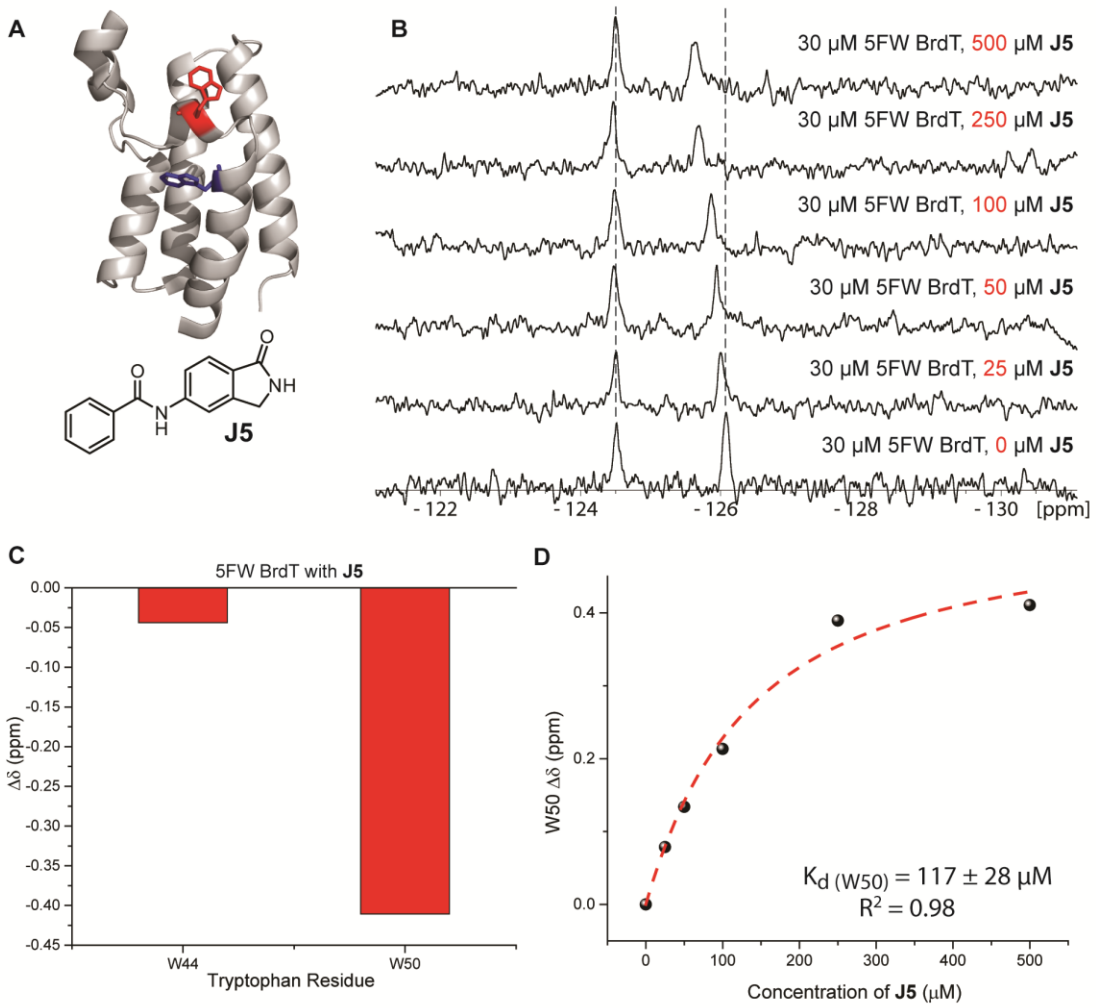


Figure 5.29 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of J5.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **J5**. C) Chemical shift perturbations for all 5FW BrdT resonances at 500 μM **J5**. D) Binding isotherm of W50 perturbations for the titration of **J5**

Table 5.27 5FW BrdT ^{19}F chemical shift perturbations at varying J5 concentrations

[J5]	Chemical shift (ppm)	W44	W50
0	δ	-124.5157	-126.0800
25	Δδ	-0.0055	-0.0787
50	Δδ	-0.0399	-0.1339
100	Δδ	-0.0283	-0.2133
250	Δδ	-0.0462	-0.3894
500	Δδ	-0.0439	-0.4107

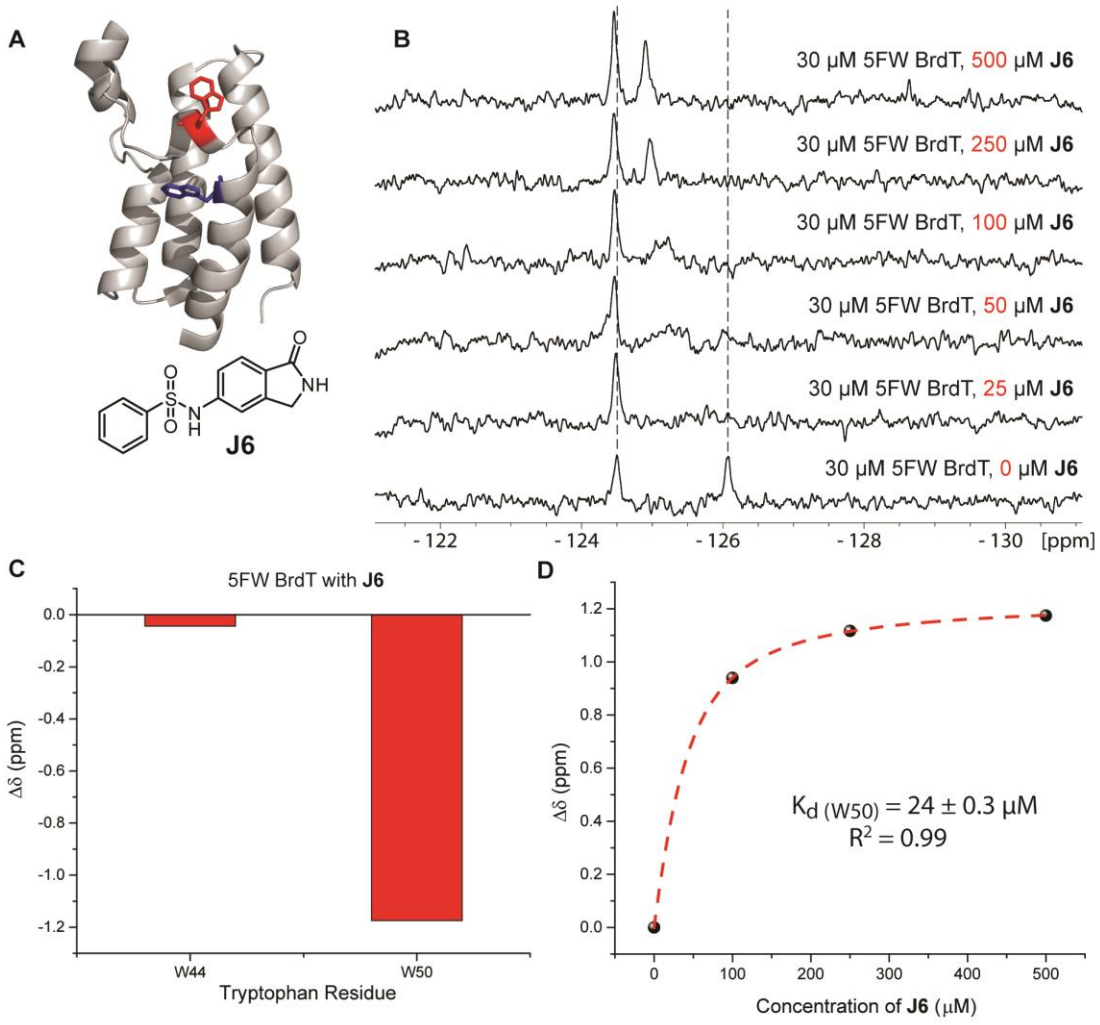


Figure 5.30 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of J6.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **J6**. C) Chemical shift perturbations for all 5FW BrdT resonances at 500 μM **J6**. D) Binding isotherm of W50 perturbations for the titration of **J6**

Table 5.28 5FW BrdT ^{19}F chemical shift perturbations at varying J6 concentrations

[J6]	Chemical shift (ppm)	W44	W50
0	δ	-124.5034	-126.0800
25	Δδ	-0.0178	
50	Δδ	-0.0336	
100	Δδ	-0.0326	-0.9400
250	Δδ	-0.0459	-1.1170
500	Δδ	-0.0439	-1.1746

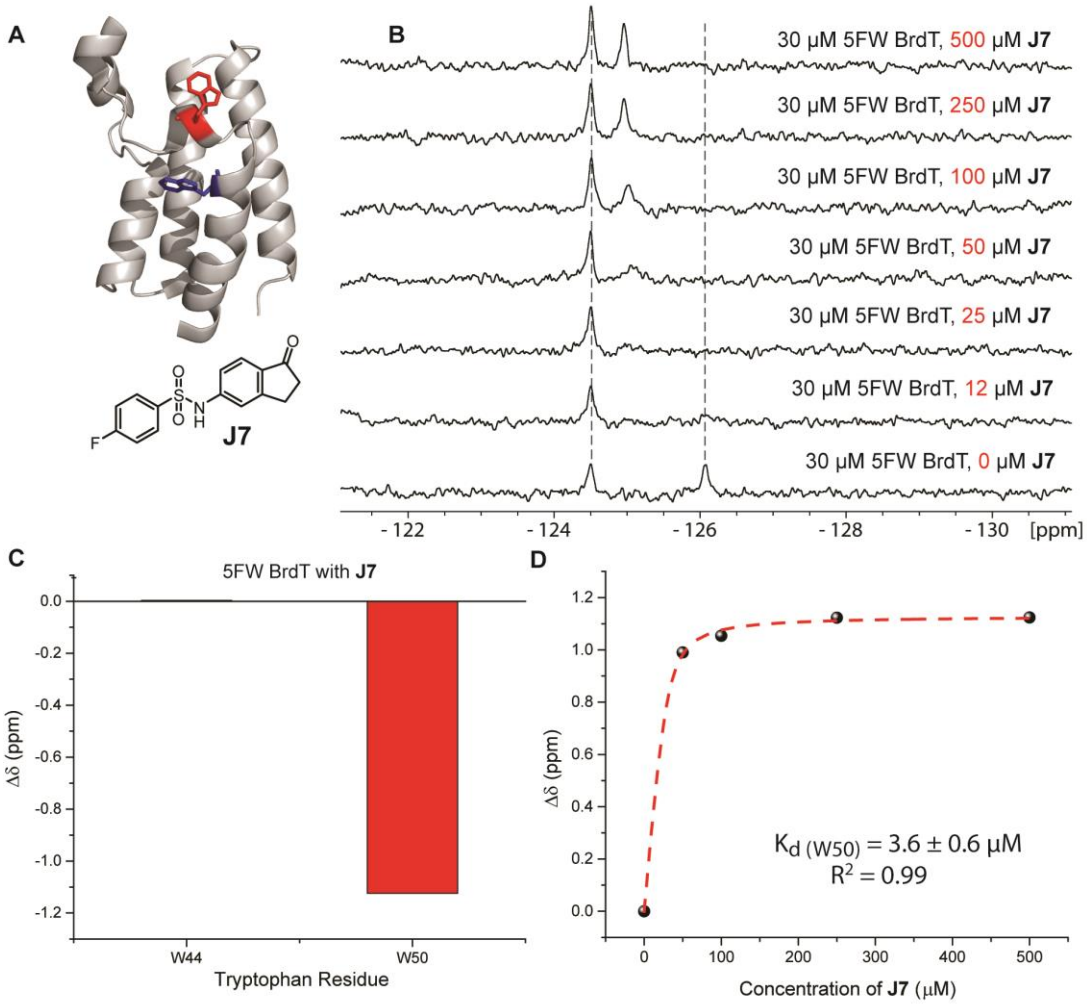


Figure 5.31 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of J7.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **J7**. C) Chemical shift perturbations for all 5FW BrdT resonances at 500 μM **J7**. D) Binding isotherm of W50 perturbations for the titration of **J7**

Table 5.29 5FW BrdT ^{19}F chemical shift perturbations at varying J7 concentrations

[J7]	Chemical shift (ppm)	W44	W50
0	δ	-124.5034	-126.0800
12	Δδ	0.0000	
25	Δδ	0.0030	
50	Δδ	-0.0016	0.9900
100	Δδ	0.0087	1.0537
250	Δδ	0.0004	1.1231
500	Δδ	0.0032	1.1237

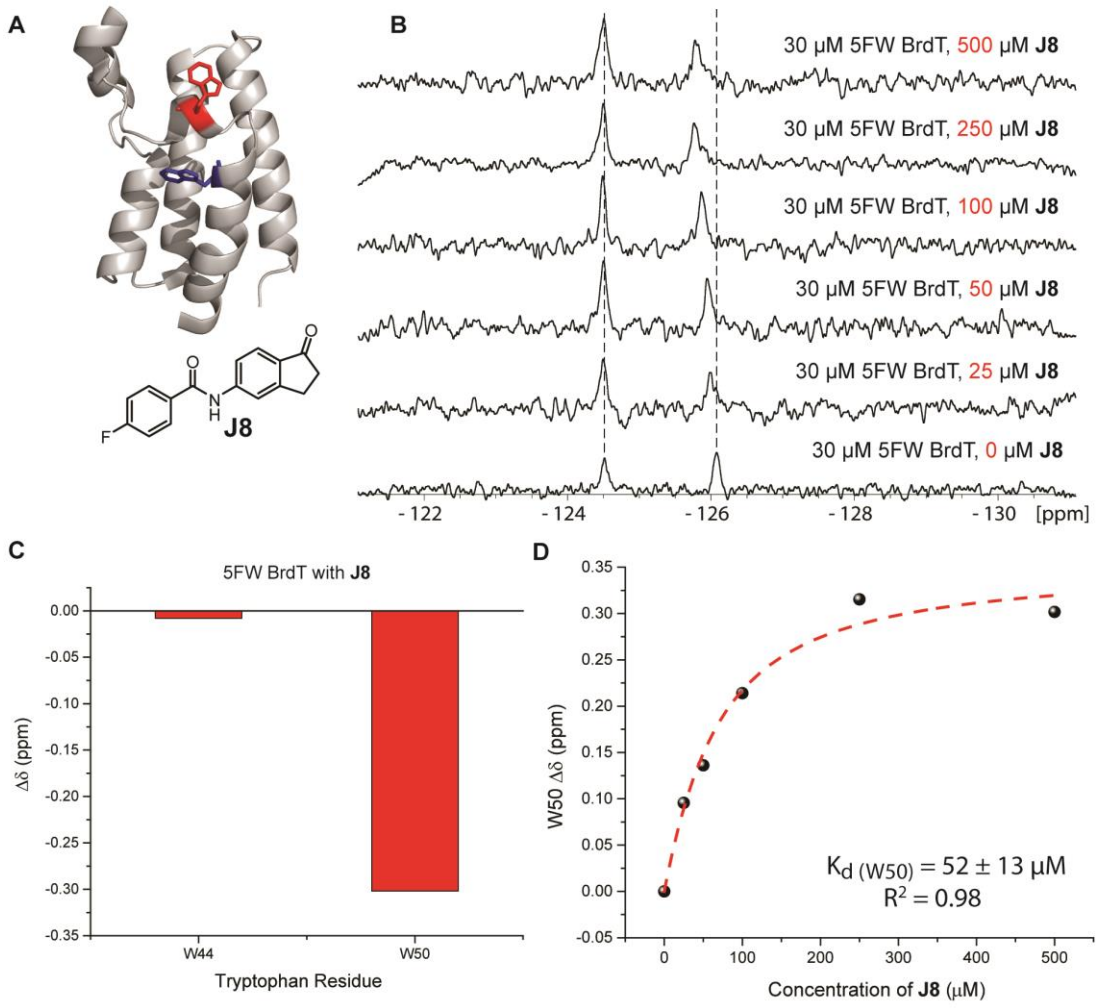


Figure 5.32 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of J8.

A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of 5FW BrdT with increasing concentrations of **J8**. C) Chemical shift perturbations for all 5FW BrdT resonances at 500 μM **J8**. D) Binding isotherm of W50 perturbations for the titration of **J8**

Table 5.30 5FW BrdT ^{19}F chemical shift perturbations at varying J8 concentrations

[J8]	Chemical shift (ppm)	W44	W50
0	δ	-124.5157	-126.0800
25	Δδ	-0.0199	-0.0956
50	Δδ	-0.0138	-0.1361
100	Δδ	-0.0248	-0.2140
250	Δδ	-0.0151	-0.3153
500	Δδ	-0.0081	-0.3017

5.4.4 Future Directions

Additional small molecule synthesis is underway to further explore these **J** analogs. Given the success of the phenyl sulfonamide class of compounds, it could be interesting to explore additional extended substitutions on the phenyl ring, where most of the current analogs are monosubstituted phenyl groups. It would also be interesting to explore other sulfonamide attachments to the indanone. Other heterocycles in place of the phenyl substitution could provide additional handles for either hydrogen bonding or reactivity. It would also be ideal to obtain crystal structure data for all of these compounds with BrdT to have a better understanding of how these compounds are oriented within the protein.

Another area that would be interesting to further study would be the different binding modes of these fragments and their impact on the chemical shifts in PrOF NMR. While most fragments only perturbed W50, some perturbed both W50 and W44 and caused them to both move downfield. Still other fragments had an effect similar to more elaborated compounds in which W44 moves downfield while W50 moves upfield. One fragment also perturbed only W44, potentially indicating a binding event that is actually taking place outside of the acetylated lysine binding pocket. It would be interesting to investigate what kinds of changes in chemical environment lead to an upfield or a downfield shift in resonances as well as how these fragments are able to induce these different effects. Our collaborators in the Schönbrunn lab may have some insight as they have crystallized a small subset of the synthesized ligands with Brd4 and BrdT.

5.5 **BET Bromodomain Inhibitors with One-Step Synthesis Discovered from Virtual Screen**

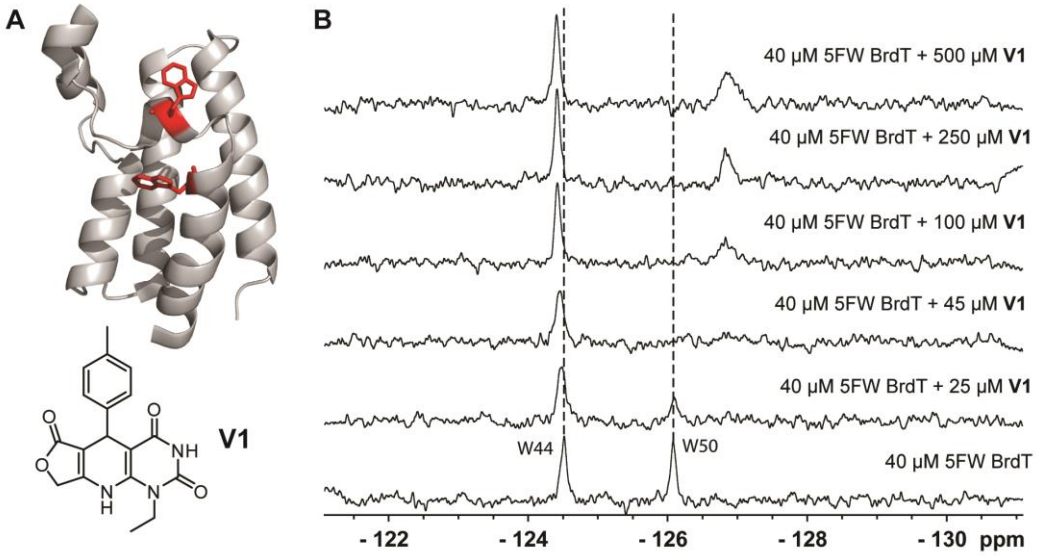


Figure 5.33 ^{19}F NMR spectral analysis of 5FW BrdT in the presence of increasing concentrations of G4.
A) Ribbon diagram of BrdT (PDB ID: 4FLP) with tryptophan side chains indicated as sticks. B) Stacked spectra of BrdT with increasing concentrations of **V1**.

To further support ligand discovery efforts, I confirmed binding of compound **V1** to BrdT via PrOF NMR (Figure 5.33). This ligand, initially discovered through the use of a virtual screen was first analyzed by FP and was reported to have an IC_{50} of 5.2 μM and a K_i of 0.69 μM for BrdT.[166] It was further characterized by DSF which reported a 4.8 °C increase in T_m for BrdT. Additionally an alpha-screen assay for **V1** against Brd4 reported an IC_{50} of 0.90 μM. PrOF NMR corroborated the results from the fluorescence polarization assay with **V1** exhibiting behavior consistent with slow to intermediate exchange (Figure 5.33B). Chemical exchange phenomenon in the slow to intermediate exchange regime, which is based on the residence time of the ligand bound to the protein and the relative frequency difference between the resonances for the bound and unbound states is consisted with a low micromolar to submicromolar binding ligand.[195, 207]

5.6 **Conclusions**

The search for a selective BrdT ligand is still ongoing. Our recent efforts have made some progress, but more work will need to be done. Currently compounds either lack the necessary potency to function as drug candidates or they are not selective enough to be used without simultaneously inhibiting the other bromodomains. Ligand deconstruction efforts on BI-2536 identified what we thought to be a reasonable selectivity hypothesis which did not hold up with the fully elaborated compound. Though unsuccessful, we learned that there must be either a different binding mode of the compound or a different protein conformational change taking place when comparing the smaller truncated analogs to the fully elaborated compound. On the fragment side, there were several core scaffolds identified that could be further optimized, including pyridones (**A** and **C**), pyridazinone (**B**), imidazoles (**D** and **E**), benzodiazepinones (**F** and **G)**, and an indanone (**J**). **J** was selected due to its optimal properties (K_d, LE, and LLE) and because the indanone scaffoled has not previously been reported in other bromodomain inhibitors, making it potentially a unique area of chemical space to explore.

5.7 **Materials, Methods, and Characterization**

5.7.1 Fluorinated Protein Expression

Fluorinated proteins were expressed as previously described.[98, 195] Briefly, BL21 *E. coli* were transformed with the plasmid of interest. Cultures were grown in nutrient rich media until they reach a high concentration at which point the nutrient rich media is exchanged for a defined media containing 5-fluoroindole (60 mg/L) for tryptophan labeling. Proteins are purified by nickel affinity chromatography, and label incorporation is verified via liquid chromatography mass spectrometry.

5.7.2 Protein Observed Fluorine NMR

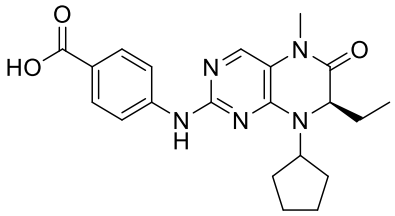
Fluorine NMR data was acquired at 470 MHz on a Bruker Avance III instrument equipped with a TCI Prodigy cryoprobe. NMR samples contained 30-50 μM 5FW BrdT or 5FW Brd4 and were prepared in a Tris buffer (50 mM Tris, 100 mM NaCl, pH 7.4)

with 5% D_2O and 50 μM trifluoroacetic acid (-76.5 ppm) as a reference. Optimized parameters for PrOF NMR experiments included a 0.6 s delay time and a 0.1 s acquisition time. Spectral width and offset were 10 ppm and -125 ppm respectively. Line broadening of 15-20 Hz was applied to all protein spectra upon data processing. Spectra were acquired at 294 K, and experiments ranged between 400 and 800 scans.

5.7.3 Small Molecule Synthesis from BI-2536 ligand deconstruction project

General reaction schemes were shown in the main text. Below is the specific procedure for each compound synthesized as well as their corresponding spectral data.

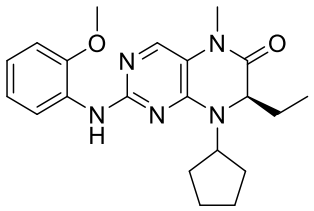
(1) (*R*)-4-((8-Cyclopentyl-7-ethyl-5-methyl-6-oxo-5,6,7,8-tetrahydropteridin-2-yl)amino)benzoic acid



G2 (100 mg, 0.338 mmol) and 4-aminobenzoic acid (72.0 mg, 0.525 mmol) were suspended in 0.3 mL of ethanol, 1.2 mL of water, and 65 μL of concentrated hydrochloric acid and refluxed for 24 h. Residual volatile compounds were evaporated under reduced pressure, and the residue was dissolved in methanol. The product precipitated out of solution upon the addition of diethyl ether. The precipitate was filtered, yielding compound **1** (31.7 mg, 23.9%)

^{1}H NMR (400 MHz, DMSO-d6) δ 7.92 (d, J = 8.9 Hz, 2H), 7.82 (s, 1H), 7.67 (d, 8.8 Hz, 2H), 4.44 (dd, 6.5, 3.0 Hz, 1H), 4.31 (quint, J = 8.9 Hz, 1H), 3.22 (s, 3H), 1.46-2.05 (m, 10H), 0.76 (t, J = 7.4, 3H)
^{13}C NMR (100 MHz, DMSO-d6) δ 166.84, 162.51, 152.16, 130.26, 119.70, 116.06, 60.46, 60.12, 28.13, 28.06, 27.86, 26.82, 22.76, 22.68, 8.26. Due to low S/N, some ^{13}C signals may be obscured in the baseline.
m/z calc'd [395.2], found [M-H] 394.2

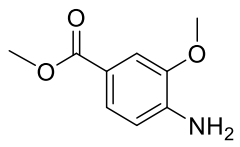
(2) (*R*)-8-Cyclopentyl-7-ethyl-2-((2-methoxyphenyl)amino)-5-methyl-7,8-dihydropteridin-6(5H)-one



G2 (100 mg, 0.338 mmol) and *o*-anisidine (62 mg, 0.509 mmol) were suspended in 0.6 mL of ethanol, 2.4 mL of water, and 65 μL of concentrated hydrochloric acid and refluxed for 24 h. Volatile compounds were evaporated under reduced pressure and redissolved in ethyl acetate. Crude compound was extracted with ethyl acetate/water three times and once with brine. The organic layer was dried over anhydrous sodium sulfite and filtered. Volatiles were evaporated under reduced pressure. The product was then purified via column chromatography and eluted near 70% EtOAc/30% Hex. Remaining solvent was evaporated under reduced pressure, yielding **2** (127.6 mg, 98.8%)

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.45 (dd, J = 7.7, 1.8 Hz, 1H), 7.67 (s, 1H), 7.45 (s, 1H), 6.94 (m, 2H), 6.88 (dd, J = 7.7, 1.7 Hz, 1H), 4.50 (quint, J = 8Hz, 1H), 4.21 (dd, J = 7.9, 3.7 Hz, 1H), 3.91 (s, 3H), 3.32 (s, 3H), 2.14 (m, 1H), 2.00 (m, 1H), 1.83 (m, 4H), 1.68 (m, 4H), 0.88 (t, J = 7.5 Hz, 3H)
^{13}C NMR (125 MHz, $CDCl_3$) δ 163.75, 155.66, 152.30, 147.62, 138.14, 129.97, 120.85, 120.66, 117.82, 115.74, 109.85, 60.00, 58.40, 55.67, 29.69, 29.19, 28.14, 27.11, 23.44, 23.06, 9.20
m/z calc'd [381.2], found [M+H] 382.2

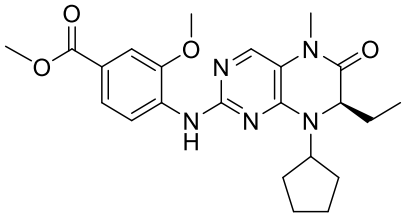
Methyl 4-Amino-3-methoxybenzoate



3-methoxy-4-aminobenzoic acid (1.05 g, 5.98 mmol) was dissolved in 100 mL of methanol. 2 mL of sulfuric acid was then added, and the solution was refluxed for 24 h. The reaction was quenched on ice with a saturated solution of sodium bicarbonate. Volatiles were evaporated under reduced pressure. The remaining residue was rinsed with ethyl acetate and water three times. The organic layer was dried over anhydrous sodium sulfite and filtered. All organic layers were combined, and volatiles were evaporated under reduced pressure to yield an orange-brown solid (1.10 g, 97%). The product was used without further purification.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.58 (dd, J = 8.1, 1.7 Hz, 1H), 7.49 (d, J = 1.6 Hz, 1H), 6.75 (d, J = 8.1 Hz), 3.92 (s, 3H), 3.89 (s, 3H)
^{13}C NMR (125 MHz, $CDCl_3$) δ 167.24, 146.52, 140.12, 123.98, 120.29, 113.82, 111.27, 55.66, 51.86

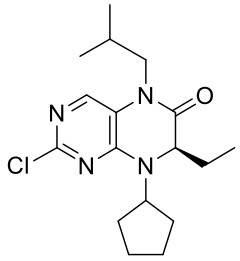
(3) Methyl (*R*)-4-((8-Cyclopentyl-7-ethyl-5-methyl-6-oxo-5,6,7,8-tetrahydropteridin-2-yl)amino)-3-methoxybenzoate



G2 (100 mg, 0.338 mmol) and methyl 4-amino-3-methoxybenzoate (95 mg, 0.509 mmol) were suspended in 0.6 ml of ethanol, 2.4 mL of water, and 0.65 μL of concentrated hydrochloric acid and refluxed for 24 h. The reaction was quenched with a saturated sodium bicarbonate solution. Volatiles were evaporated under reduced pressure. The residue was washed three times with ethyl acetate and water, and once with brine. Organic layers were combined and dried over anhydrous sodium sulfite. The crude product was then purified via column chromatography and eluted between ~60%-80% EtOAc in hexanes. Volatiles were evaporated under reduced pressure, yielding **3** (41.5 mg, 27.9%)

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.59 (d, J = 8.6 Hz, 2H), 7.70 (dd, J = 8.5, 1.8 Hz, 1H), 7.54 (d, J = 1.7 Hz, 1H), 4.52 (quint, J = 8.5 Hz, 1H), 4.24 (dd, J = 7.9, 3.6 Hz, 1H), 3.99 (s, 3H), 3.92 (s, 3H), 3.34 (s, 3H), 2.17 (m, 1H), 2.01 (m, 1H), 1.85 (m, 4H), 1.72 (m, 4H), 0.89 (t, J = 7.5 Hz, 3H)
^{13}C NMR (125 MHz, $CDCl_3$) δ 167.17, 163.76, 154.97, 152.27, 146.64, 123.44, 121.67, 115.76, 110.39, 59.92, 58.47, 55.92, 51.90, 29.73, 29.21, 28.18, 27.19, 23.50, 23.10, 9.18
m/z calc'd [439.2], found [M+H} 440.3

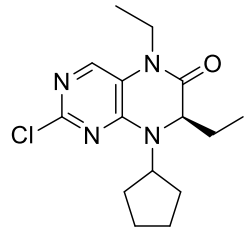
(4) (*R*)-2-Chloro-8-cyclopentyl-7-ethyl-5-isobutyl-7,8-dihydropteridin-6(5H)-one



Sodium hydride in a 60% dispersion in mineral oil (36 mg, 0.95 mmol) was dissolved in 2 mL of dimethylacetamide (DMA) and stirred under an N_2 atmosphere. **G1** (204 mg, 0.727 mmol) was dissolved in minimal DMA and added dropwise to stirred NaH solution on ice. The reaction was stirred at 0 °C for 90 min. Isobutyl iodide (108 µL, 0.95 mmol) was added, and the reaction continued to stir for 21 hours while gradually warming to room temperature. The reaction was quenched by the addition of crushed ice and 4 mL of sat. ammonium chloride. The solution was then extracted three times with ethyl acetate/water. The organic layer was dried over anhydrous sodium sulfite, and volatiles were evaporated under reduced pressure to yield a viscous yellow oil. The crude product was then purified via column chromatography to yield product **4** (119 mg, 48 %) which eluted at approximately ~20%EtOAc/80%Hex.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.69 (s, 1H), 4.40 (quint, J = 8.7 Hz, 1H), 4.20 (dd, J = 8.3, 3.6 Hz, 1H), 3.84 (dd, J = 14.0, 7.0 Hz, 1H), 3.57 (dd, J = 14.1, 7.8 Hz, 1H), 1.58-2.14 (m, 11H), 0.96 (d, J = 6.7 Hz, 3H), 0.94 (d, J = 6.7 Hz, 3H), 0.90 (t, J = 7.7 Hz)
^{13}C NMR (125 MHz, $CDCl_3$) δ 163.73, 153.48, 152.50, 138.18, 120.67, 60.62, 59.59, 48.65, 29.47, 29.40, 27.14, 26.48, 23.93, 23.66, 20.15, 19.99, 9.21
m/z calc'd [336.2], found, [M+H] 337.1

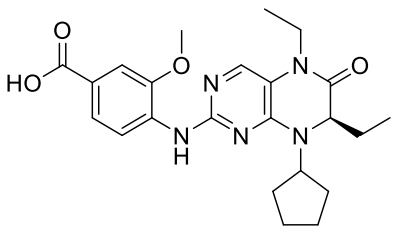
(**G3)** (*R*)-2-Chloro-8-cyclopentyl-5,7-diethyl-7,8-dihydropteridin-6(5H)-one



Sodium hydride in a 60% dispersion in mineral oil (55 mg, 1.3 mmol) was dissolved in 2 mL of dimethylacetamide (DMA) and stirred under an N_2 atmosphere. **G1** (308 mg, 1.1 mmol) was dissolved in minimal DMA and added dropwise to stirred NaH solution on ice. The reaction was stirred at 0 °C for 1H. Iodoethane (222 mg, 1.4 mmol) was added, and the reaction continued to stir for 18 hours while gradually warming to room temperature. The reaction was quenched by the addition of crushed ice and 5 mL of sat. ammonium chloride. The solution was then extracted three times with ethyl acetate/water. The organic layer was dried over sodium sulfite, and volatiles were evaporated under reduced pressure to yield **G3** as a viscous oil (208 mg, 61.6 %). The material was carried on without further purification.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.70 (s, 1H), 4.34 (quint, J = 8.4 Hz, 1H), 4.23 (dd, J = 7.4, 3.5 Hz, 1H), 4.09 (dq, J = 14.6, 7.5 Hz, 1H), 3.76 (dq, J = 14.5, 7.3 Hz, 1H), 1.55-2.22 (m, 10H), 1.26 (t, J = 7.4 Hz, 3H), 0.86 (t, J = 7.5 Hz, 3H)
^{13}C NMR (125 MHz, $CDCl_3$) δ 163.27, 153.56, 152.59, 137.41, 119.83, 60.57, 59.48, 36.61, 29.42, 29.22, 27.26, 23.96, 23.78, 12.07, 8.8.
m/z calc'd [308.1], found [M+H] 309.1

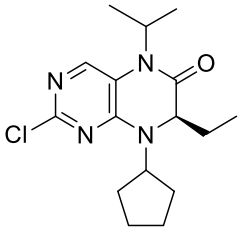
(5) (*R*)-4-((8-Cyclopentyl-5,7-diethyl-6-oxo-5,6,7,8-tetrahydropteridin-2-yl)amino)-3-methoxybenzoic acid



G3 (92.2 mg, 0.209 mmol) and 3-methoxy-4-aminobenzoic acid (55 mg, 0.315 mmol) were suspended in 0.6 mL of ethanol, 2.4 mL of water, and 65 μL of concentrated hydrochloric acid and refluxed for 48 h. Residual volatile compounds were evaporated under reduced pressure, and the residue was dissolved in methanol. The crude product was purified via column chromatography and eluted between 80% and 100% EtOAc in hexanes. Volatiles were evaporated under reduced pressure to yield **5** as a viscous oil. (10.8 mg, 11.7%)

^{1}H NMR (500 MHz, $CDCl_3$) δ 8.62 (d, J = 8.5 Hz, 1H), 7.85 (s, 1H), 7.80 (d, J = 8.6 Hz, 1H), 7.75 (s, 1H), 7.61 (s, 1H), 4.55 (quint, J = 7.7 Hz, 1H), 4.25 (dd, J = 7.7, 3.5 Hz, 1H), 4.11 (m, 1H), 4.01 (s, 1H), 3.80 (m, 1H), 2.19 (m, 1H), 2.03 (m, 1H), 1.86 (m, 4H), 1.73 (m, 4H), 1.30 (dd, J = 13, 5 Hz, 3H), 0.90 (t, J = 7.3 Hz, 3H)
^{13}C NMR (125 MHz, $CDCl_3$) δ 163.25, 154.70, 152.61, 146.77, 137.35, 135.23, 124.28, 121.00, 115.91, 115.07, 110.83, 59.72, 58.53, 55.95, 36.45, 29.78, 29.70, 29.20, 27.10, 23.50, 23.11, 12.38,
m/z calc'd [439.2], found [M+H] 440.3

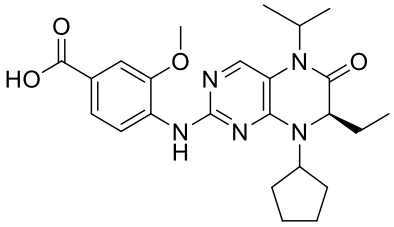
(G4) (*R*)-2-Chloro-8-cyclopentyl-7-ethyl-5-isopropyl-7,8-dihydropteridin-6(5H)-one



Sodium hydride in a 60% dispersion in mineral oil (47 mg, 1.18 mmol) was dissolved in 2 mL of DMA and stirred under an N_2 atmosphere. **G1** (255 mg, 0.9 mmol) was dissolved in minimal DMA and added dropwise to stirred NaH solution on ice. Reaction was stirred at 0 °C for 1H. Isopropyl iodide (108 μL, 1.18 mmol) was added, and the reaction continued to stir for 22 hours while gradually warming to room temperature. Reaction was quenched by the addition of crushed ice and 5 mL of sat. ammonium chloride. Solution was then extracted three times with ethyl acetate/water. Organic layer was dried over sodium sulfite, and volatiles were evaporated under reduced pressure. The crude product was then purified via column chromatography and eluted at approximately ~15%EtOAc/80%Hex. Volatiles were again evaporated under reduced pressure to afford **G4** (79.2 mg, 27.2%)

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.89 (s, 1H), 4.88 (m, 1H), 4.35 (quint, J = 8.3 Hz, 1H), 4.12 (m, 1H, obstructed by impurity), 1.57-2.10 (m, 10H), 1.50 (d, J = 7.1 Hz, 3H), 1.43 (d, J = 7.1 Hz, 3H), 0.85 (t, J = 7.5 Hz, 3H)
^{13}C NMR (125 MHz, $CDCl_3$) δ 164.60, 153.46, 153.38, 139.16, 119.63, 60.78, 59.04, 46.67, 29.61, 29.43, 26.52, 23.91, 23.66, 20.48, 18.53, 9.10
m/z calc'd [322.15], found, [M+Na] 345.1

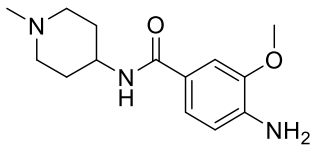
(6) (*R*)-4-((8-Cyclopentyl-7-ethyl-5-isopropyl-6-oxo-5,6,7,8-tetrahydropteridin-2-yl)amino)-3-methoxybenzoic acid



G4 (35.7 mg, 0.11 mmol) and 3-methoxy-4-aminobenzoic acid (27.6 mg, 0.165 mmol) were suspended in 0.6 mL of ethanol, 2.4 mL of water, and 25 μL of concentrated hydrochloric acid and refluxed for 72 h. Residual volatile compounds were evaporated under reduced pressure, and the residue was dissolved in methanol. The crude product was purified via column chromatography and eluted near 10% MeOH/90% dichloromethane. Volatiles were evaporated under reduced pressure to yield **6** as a white solid. (14.35 mg, 28.7%)

^{1}H NMR (500 MHz, MeOD) δ 8.65 (d, J = 8.7 Hz, 1H), 7.93 (s, 1H), 7.84 (s, 1H), 7.76 (dd, J = 8.6, 1.8 Hz, 1H), 7.59 (d, J = 1.7 Hz, 1H), 4.85 (m, 1H), 4.51 (m, 1H), 4.09 (m, 1H), 4.00 (s, 3H), 1.58-2.22 (m, 10H), 1.51 (d, J = 7.1 Hz, 3H), 1.44 (d, J = 7.1 Hz, 3H), 0.86 (t, J = 7.5 Hz, 3H)
^{13}C NMR (125 MHz, MeOD) δ 170.50, 164.56, 154.39, 153.49, 146.89, 138.93, 135.52, 124.14, 120.75, 115.87, 115.32, 110.78, 60.36, 58.40, 56.04, 46.56, 29.90, 29.55, 26.16, 23.63, 23.20, 20.49, 18.56, 9.24
m/z calc'd [453.2], found [M-H] 452.2

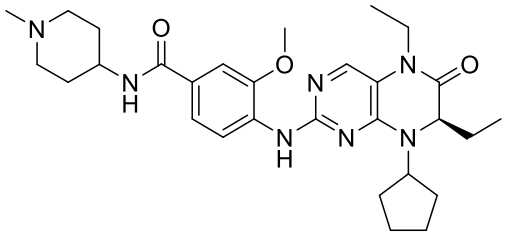
4-Amino-3-methoxy-N-(1-methylpiperidin-4-yl)benzamide



300 mg, 1.80 mmol of 3-methoxy-4-amino benzoic acid dissolved in 3 mL DMF. EDCI (335 mg, 2.16 mmol) was added along with 22 mg of DMAP. Once the material dissolved, 0.225 mL, 1.80 mmol of methyl amino piperidine was added. The reaction was left for 4 days. 1 eq. of triethylamine was added, and the reaction was left to stir for an additional day. Solid material was filtered, rinsed with DCM and acetonitrile. Volatiles were evaporated under reduced pressure to form a viscous oil. The sample was left at room temperature, and the product recrystallized from the solution. (87.7 mg, 18.5% yield). This material was used without further purification.

^{1}H NMR (500 MHz, MeOD) δ 7.36 (d, 1.9 Hz, 1H), 7.32 (dd, J = 8.1, 2.0 Hz, 1H), 6.72 (d, J = 8.2 Hz, 1H), 3.92 (m, 4H), 2.95 (d, J = 12.2 Hz, 2H), 2.33 (s, 3H), 2.21 (td, J = 12.3, 2.1 Hz, 2H), 1.96 (dd, J = 12.5, 2.2 Hz, 2H), 1.70 (qd, J = 12.5, 4.0 Hz, 2H)
^{13}C NMR (125 MHz, MeOD) δ 168.57, 146.42, 141.03, 122.50, 120.90, 112.85, 109.13, 54.65, 54.37, 46.62, 44.73, 30.97
m/z calc'd [263.2], found [M+H] 264.2

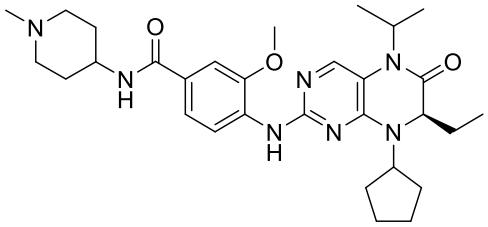
BI-2536-Et (*R*)-4-((8-Cyclopentyl-5,7-diethyl-6-oxo-5,6,7,8-tetrahydropteridin-2-yl)amino)-3-methoxy-N-(1-methylpiperidin-4-yl)benzamide



G3 (23.8 mg, 0.077 mmol) and 4-amino-3-methoxy-N-(1-methylpiperidin-4-yl)benzamide (30 mg, 0.165 mmol) were suspended in 0.3 mL of ethanol, 1.2 mL of water, and 65 μL of concentrated hydrochloric acid and refluxed for 72 h. Residual volatile compounds were evaporated under reduced pressure, and the residue was dissolved in methanol. The crude product was purified via column chromatography (DCM/MeOH). Volatiles were evaporated under reduced pressure to yield an oil. (14.28 mg, 34%)

^{1}H NMR (500 MHz, MeOD) δ 8.51 (d, J = 9.0 Hz, 1H), 7.83 (s, 1H), 7.50-7.55 (m, 2H), 4.54 (quint, J = 8.4 Hz, 1H), 4.26 (dd, J = 7.8, 3.6 Hz,f 1H), 4.02-4.14 (m, 1H), 4.02 (s, 3H), 3.78-3.90 (m, 1H), 2.83 (t, J = 12.4 Hz, 2H), 2.69 (s, 3H), 1.65-2.22 (m, 16H), 1.25 (t, J = 7.1 Hz, 3H), 0.88 (t, J = 7.6 Hz, 3H)
^{13}C NMR (125 MHz, MeOD) δ 168.04. 163.67. 154.87*, 152.66. 147.20. 137.75. 133.22, 126.00, 120.06, 116.10, 114.49*, 108.75, 59.78, 58.83, 55.22, 53.63, 45.36, 43.19, 36.14, 29.52, 29.10, 28.59, 26.48, 22.98, 22.66, 11.17, 7.88. *Chemical shifts are approximated due to low intensity
m/z calc'd [535.3], found [M+D] 537.3

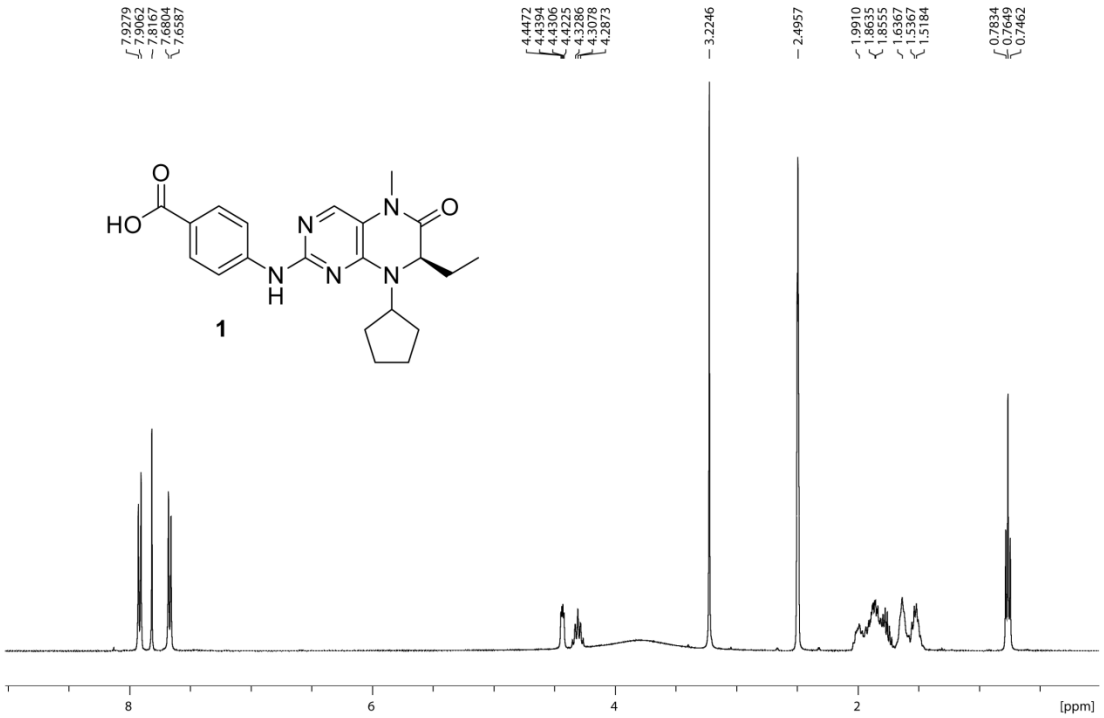
BI-2536-iPr (*R*)-4-((8-Cyclopentyl-7-ethyl-5-isopropyl-6-oxo-5,6,7,8-tetrahydropteridin-2-yl)amino)-3-methoxy-N-(1-methylpiperidin-4-yl)benzamide



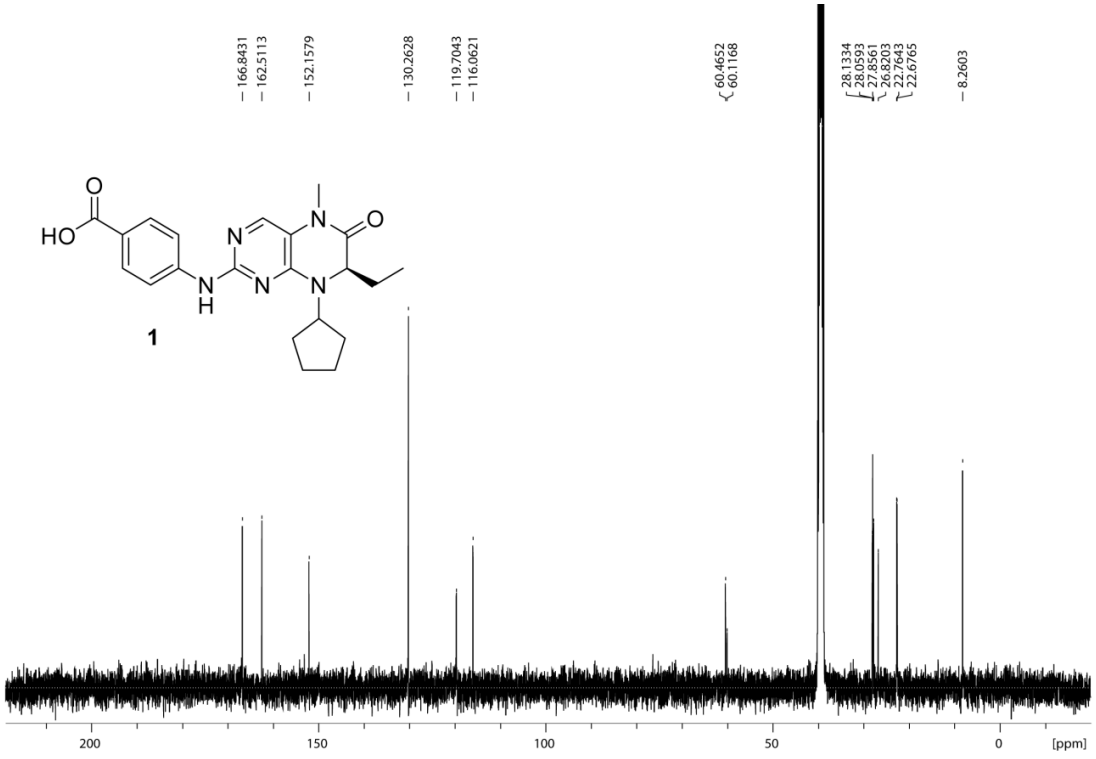
G4 (32.6 mg, 0.101 mmol) and 4-amino-3-methoxy-N-(1-methylpiperidin-4-yl)benzamide (30 mg, 0.165 mmol) were suspended in 0.6 mL of ethanol, 2.4 mL of water, and 25 μL of concentrated hydrochloric acid and refluxed for 72 h. Residual volatile compounds were evaporated under reduced pressure, and the residue was dissolved in methanol. The crude product was purified via column chromatography (DCM/MeOH). Volatiles were evaporated under reduced pressure to yield a white solid. (28.53 mg, 51.5%)

^{1}H NMR (500 MHz, MeOD) δ 8.52 (d, J = 8.9 Hz, 1H), 8.00 (s, 1H), 7.50-7.55 (m, 2H), 4.70-4.82 (m, 1H), 4.53 (quint, J = 8.1 Hz, 1H), 4.13 (dd, J = 8.4, 4.1 Hz, 1H), 4.02 (s, 3H), 3.15 (d, J = 12.2 Hz, 2H), 2.47-2.58 (m, 5H), 1.60-2.20 (m, 16H), 1.53 (d, J = 7.1 Hz, 3H), 1.48 (d, J = 7.1 Hz, 3H), 0.88 (t, J = 7.5 Hz, 3H)
^{13}C NMR (125 MHz, MeOD) δ 167.85, 165.08, 154.69, 153.45, 147.20139.24, 132.889, 126.19, 120.03, 116.13, 114.74, 108.76, 60.27, 58.48, 55.21, 53.99, 46.06, 43.04, 30.20, 29.29, 29.02, 25.52, 23.04, 22.63, 19.19, 17.66, 8.28
m/z calc'd [549.3], found [M+H] 550.3

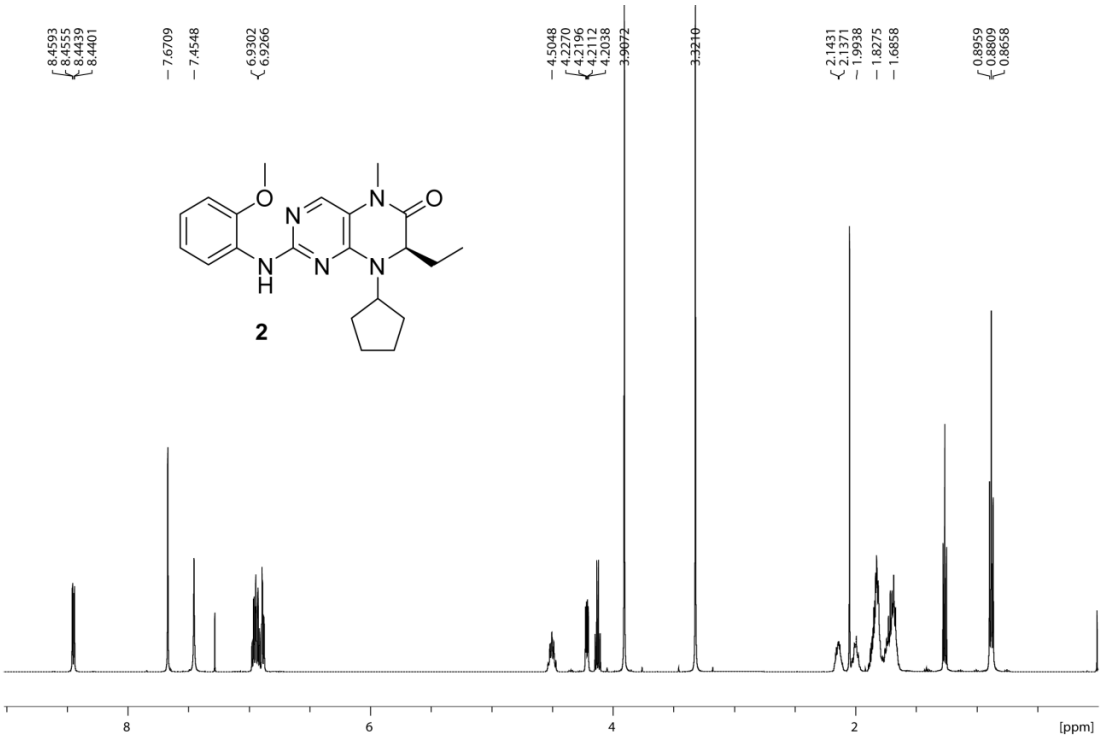
^{1}H NMR of **1** (400 MHz, DMSO-d6)



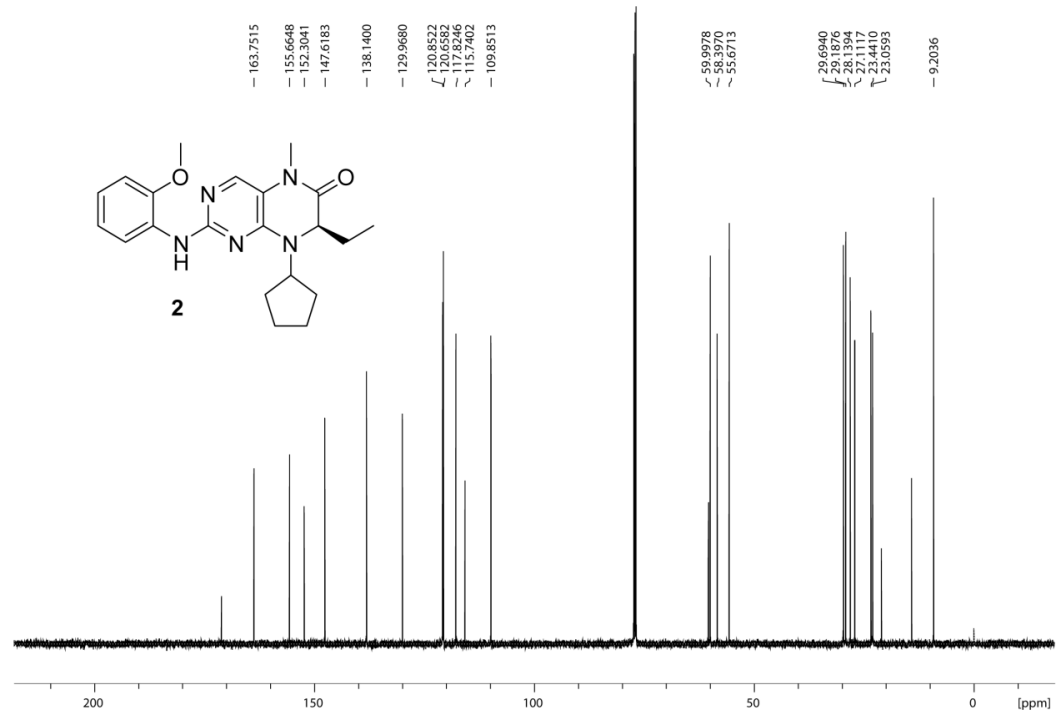
^{13}C NMR of **1** (100 MHz, DMSO-d6)



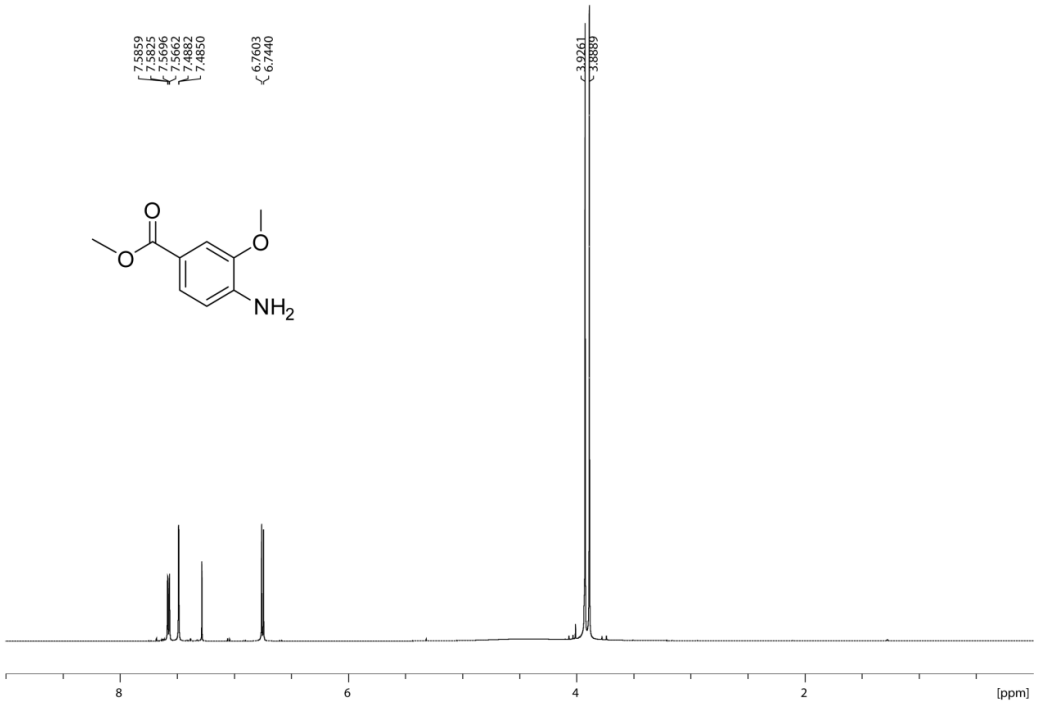
^{1}H NMR of **2** (500 MHz, $CDCl_3$)



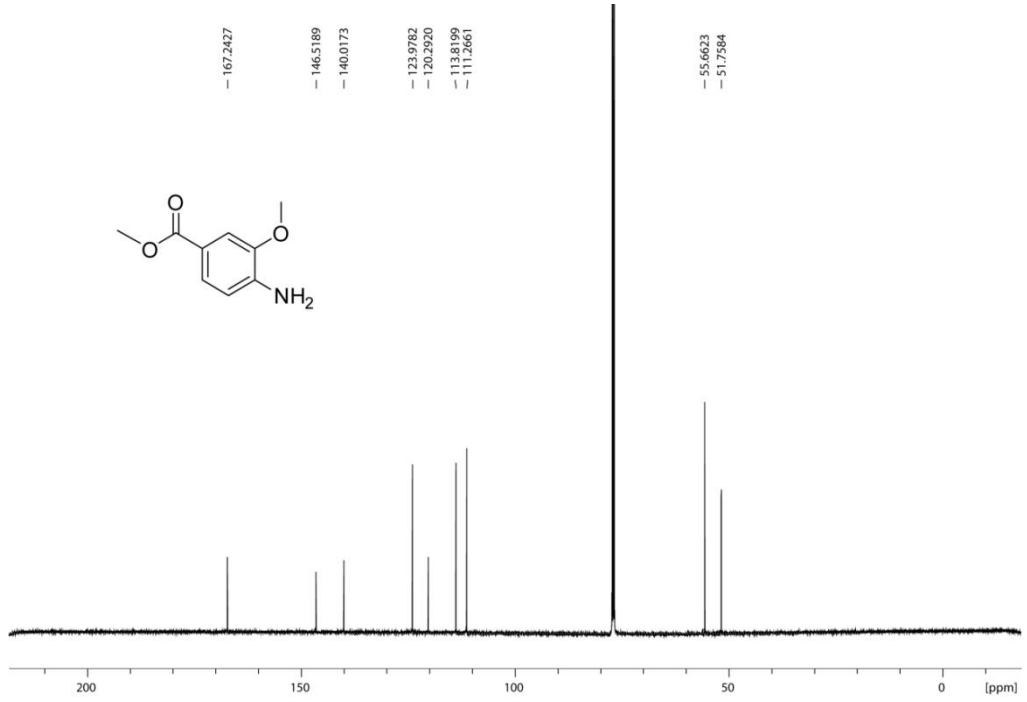
^{13}C NMR of **2** (125 MHz, $CDCl_3$)



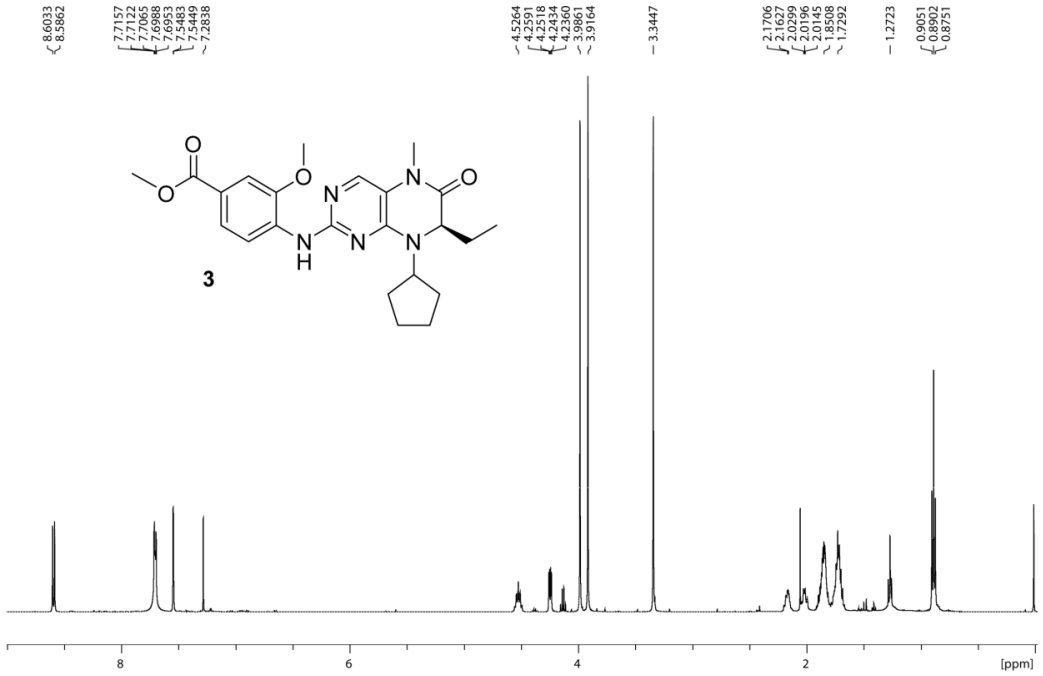
^{1}H NMR of methyl 4-amino-3-methoxybenzoate (500 MHz, $CDCl_3$)



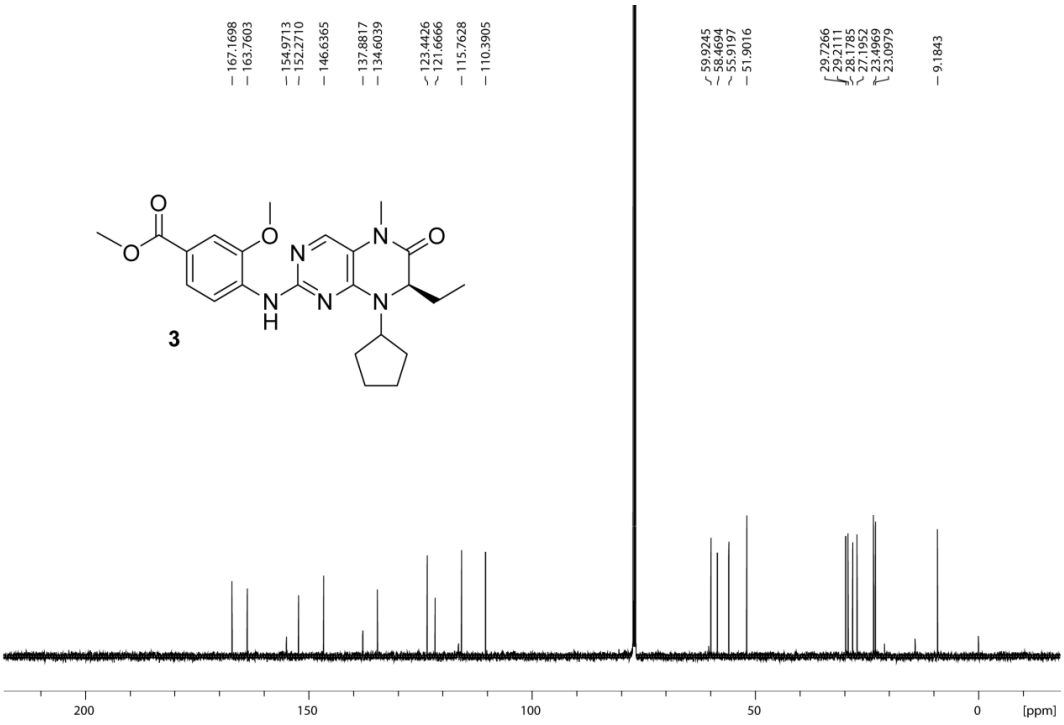
^{13}C NMR of methyl 4-amino-3-methoxybenzoate (125 MHz, $CDCl_3$)



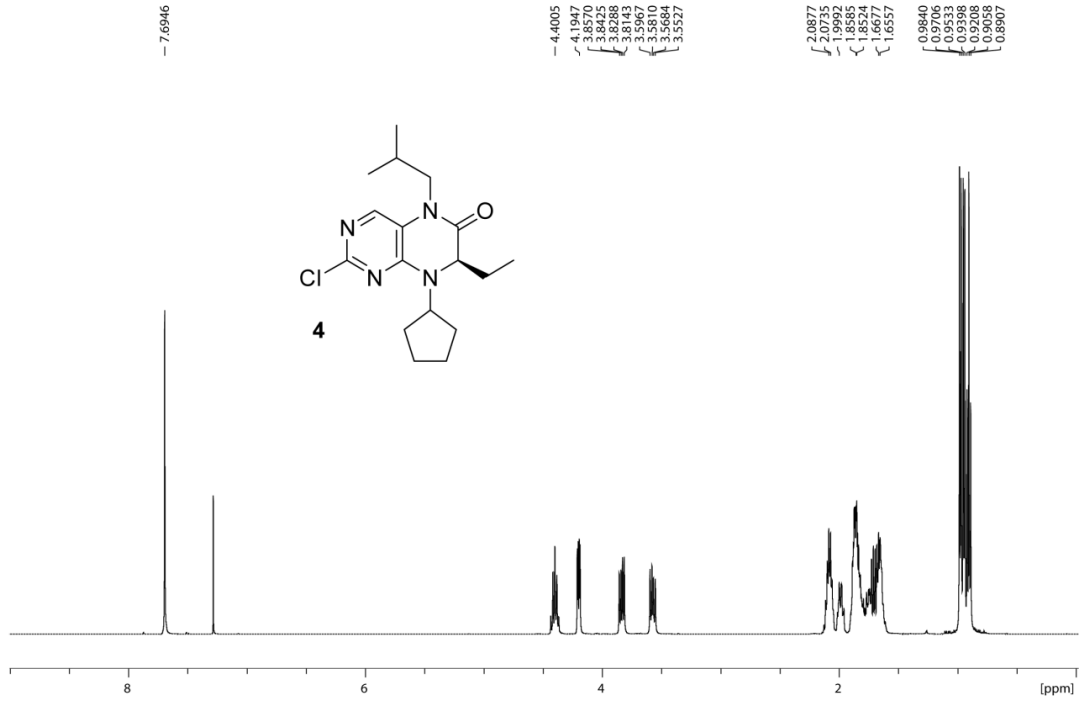
^{1}H NMR of **3** (500 MHz, $CDCl_3$)



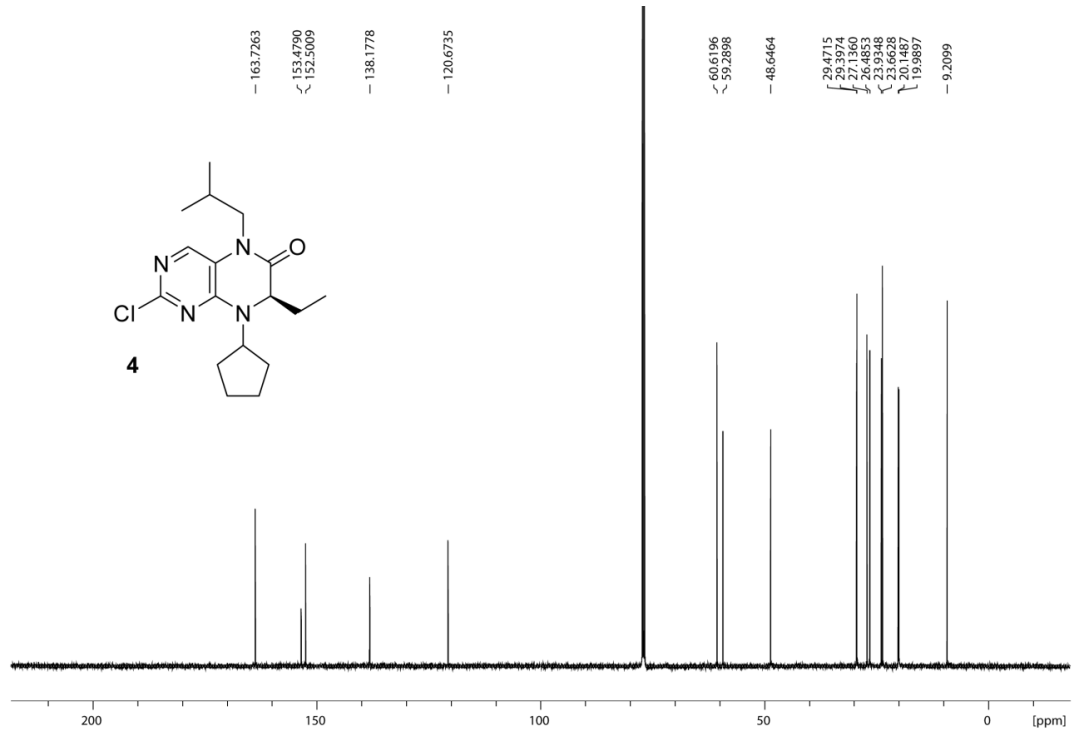
^{13}C NMR of **3** (125 MHz, $CDCl_3$)



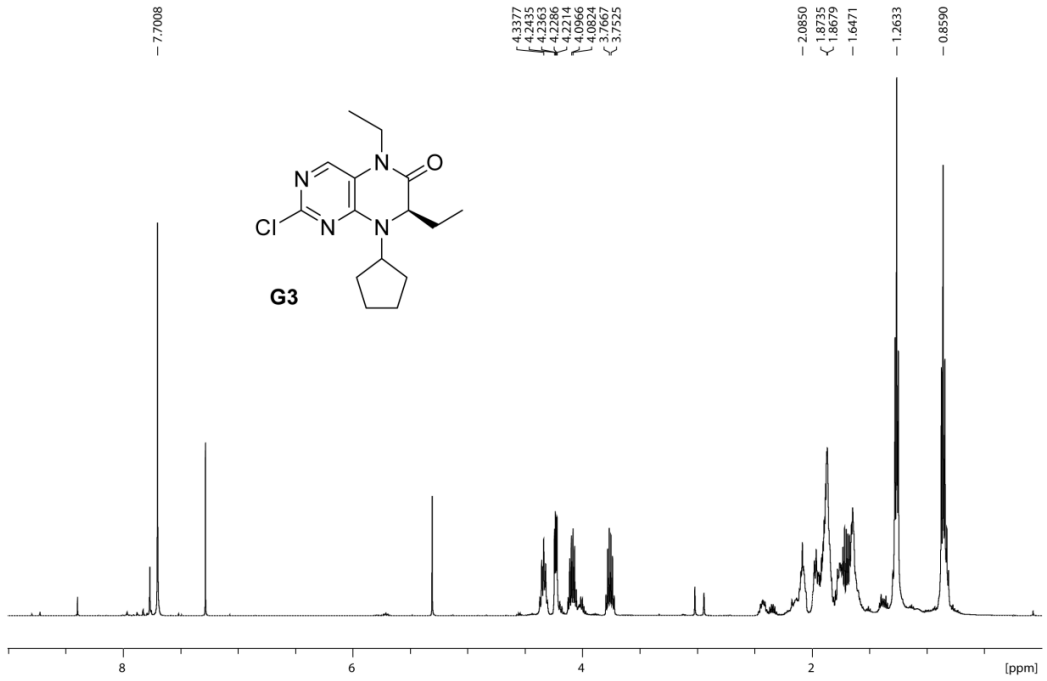
^{1}H NMR of **4** (500 MHz, $CDCl_3$)



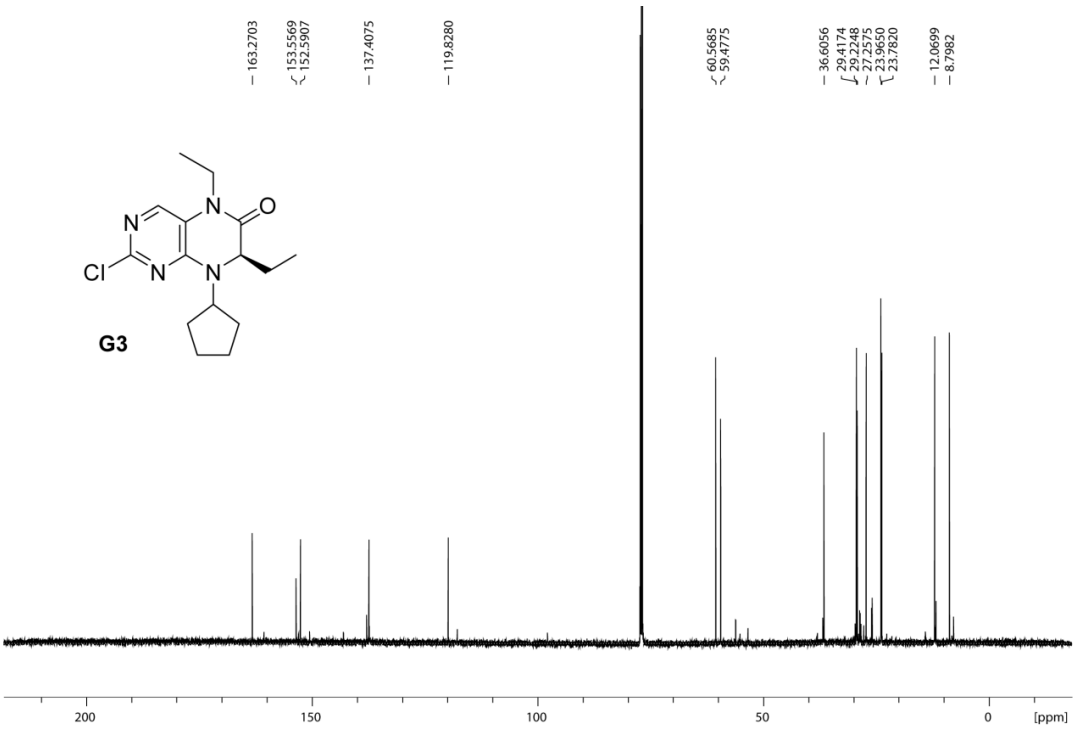
^{13}C NMR of **4** (125 MHz, $CDCl_3$)



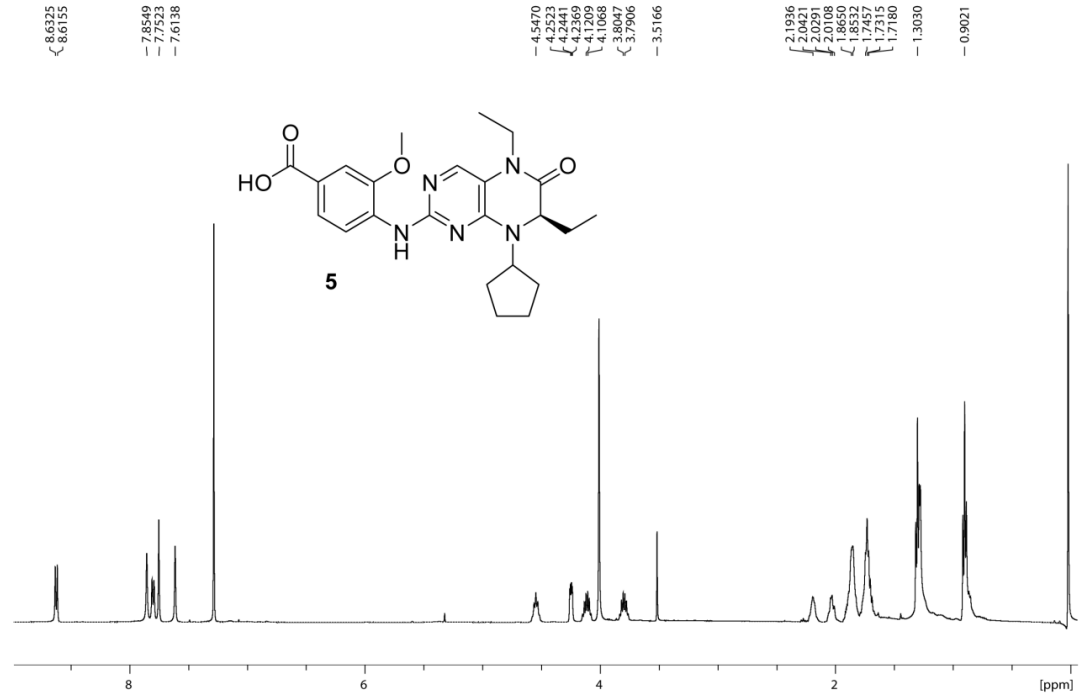
^{1}H NMR of **G3** (500 MHz, $CDCl_3$)



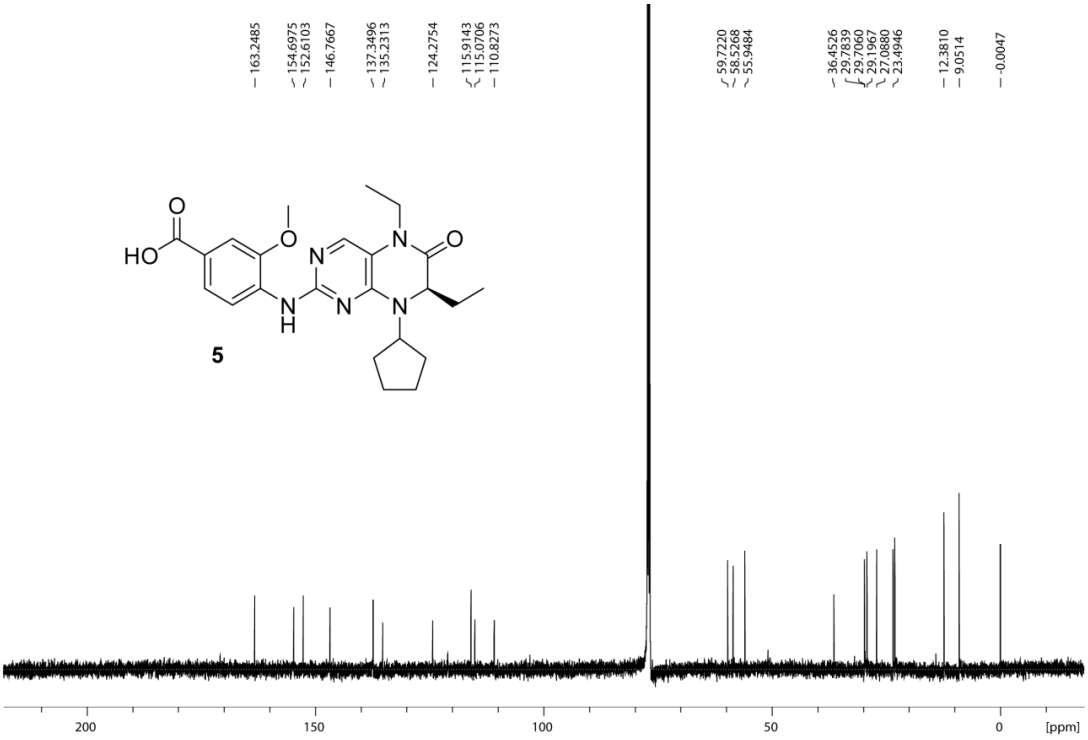
^{13}C NMR of **G3** (125 MHz, $CDCl_3$)



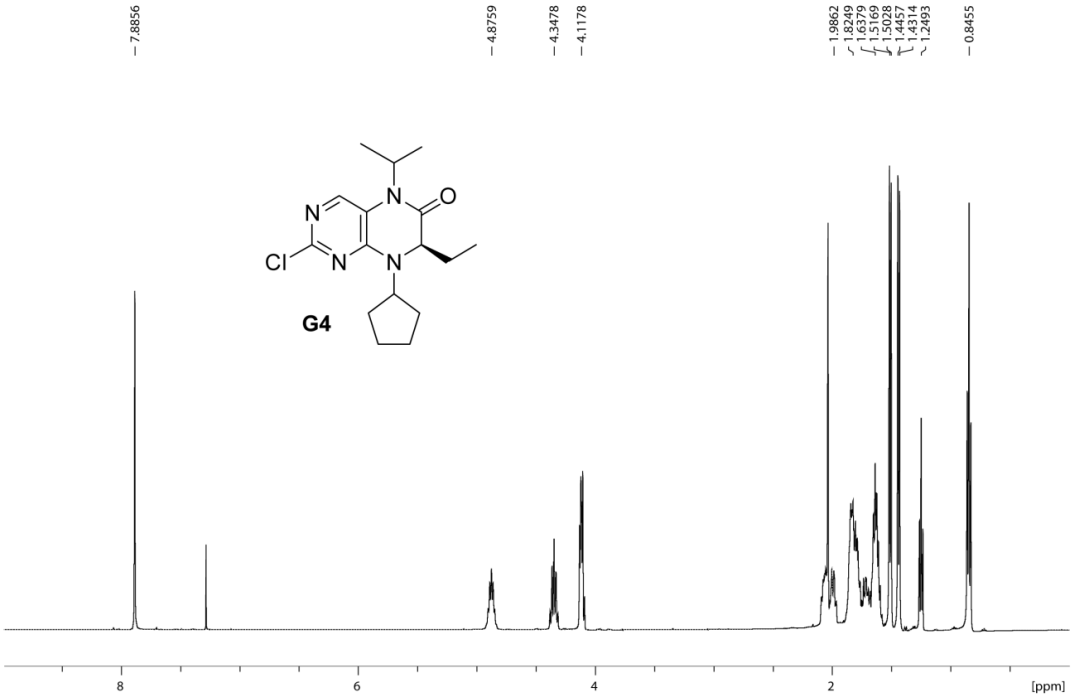
^{1}H NMR of **5** (500 MHz, $CDCl_3$)



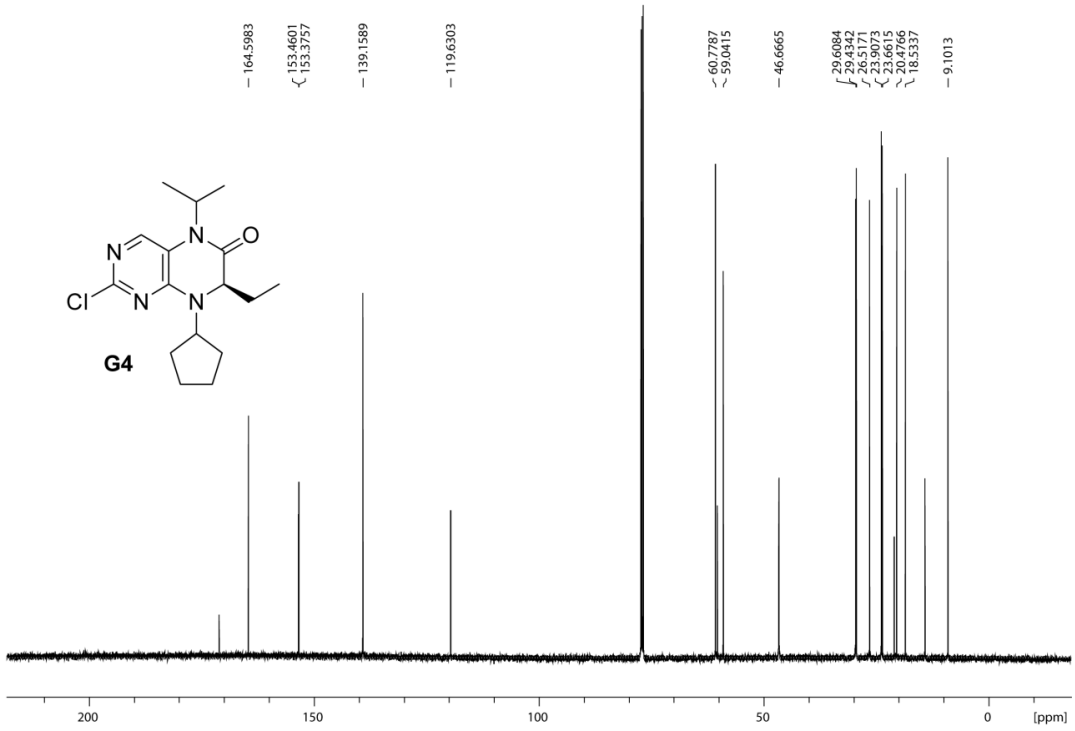
^{13}C NMR of **5** (125 MHz, $CDCl_3$)



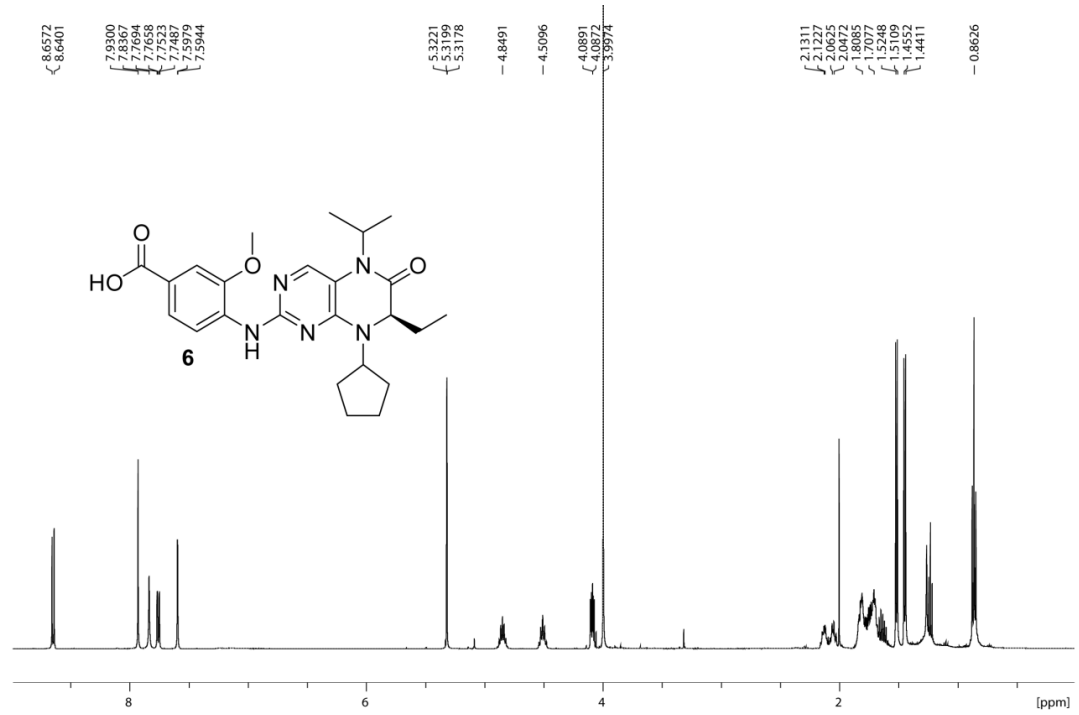
^{1}H NMR of **G4** (500 MHz, $CDCl_3$)



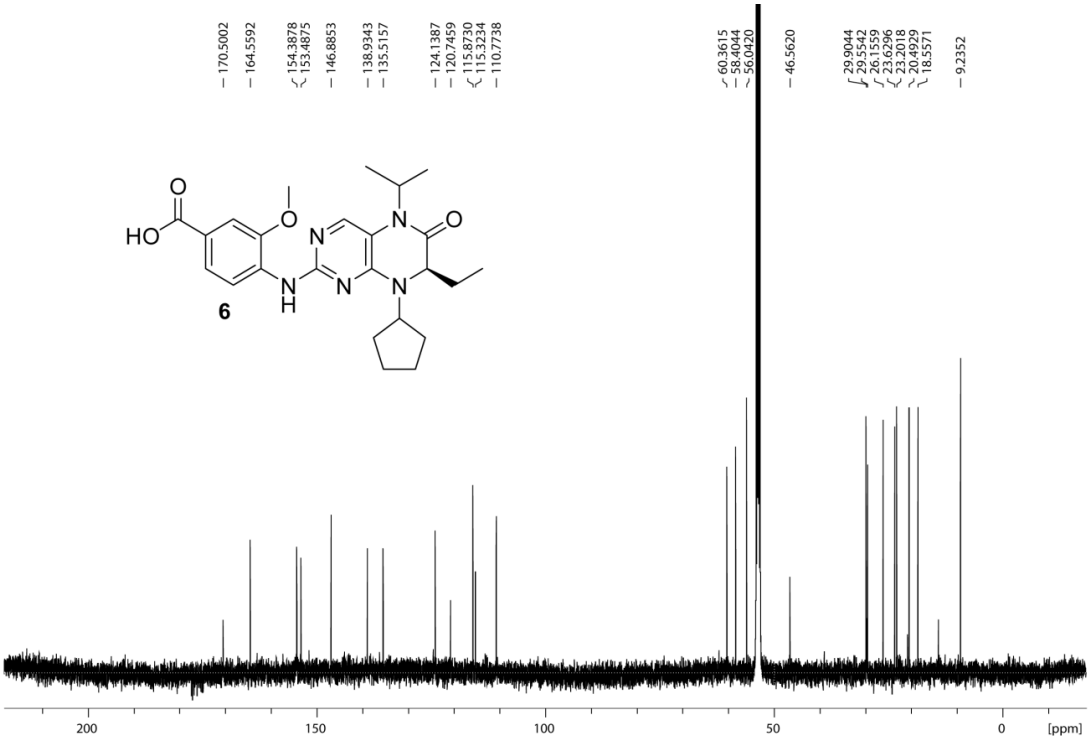
^{13}C NMR of **G4** (125 MHz, $CDCl_3$)



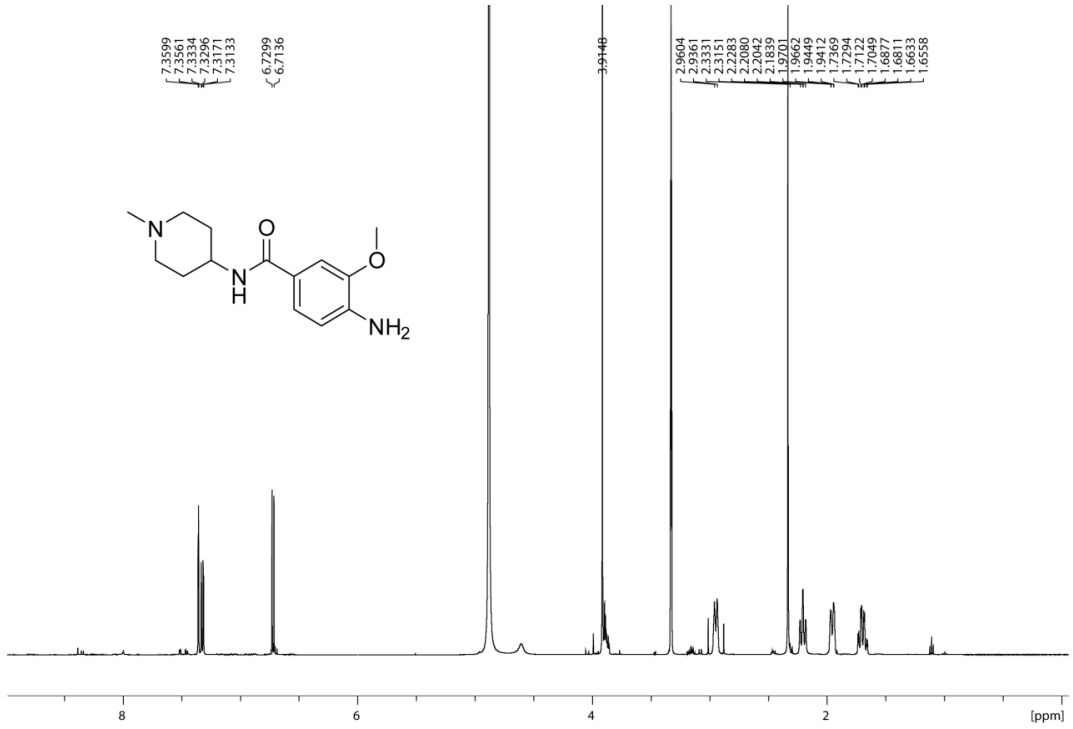
^{1}H NMR of **6** (500 MHz, MeOD)



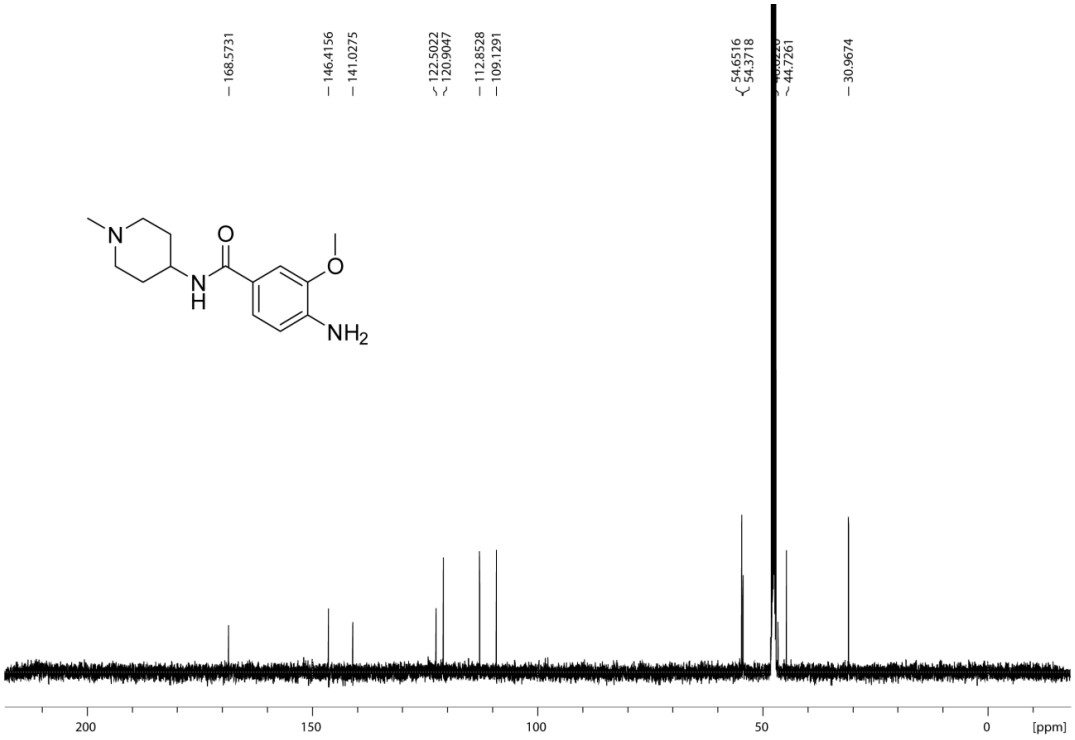
^{13}C NMR of **6** (125 MHz, MeOD)



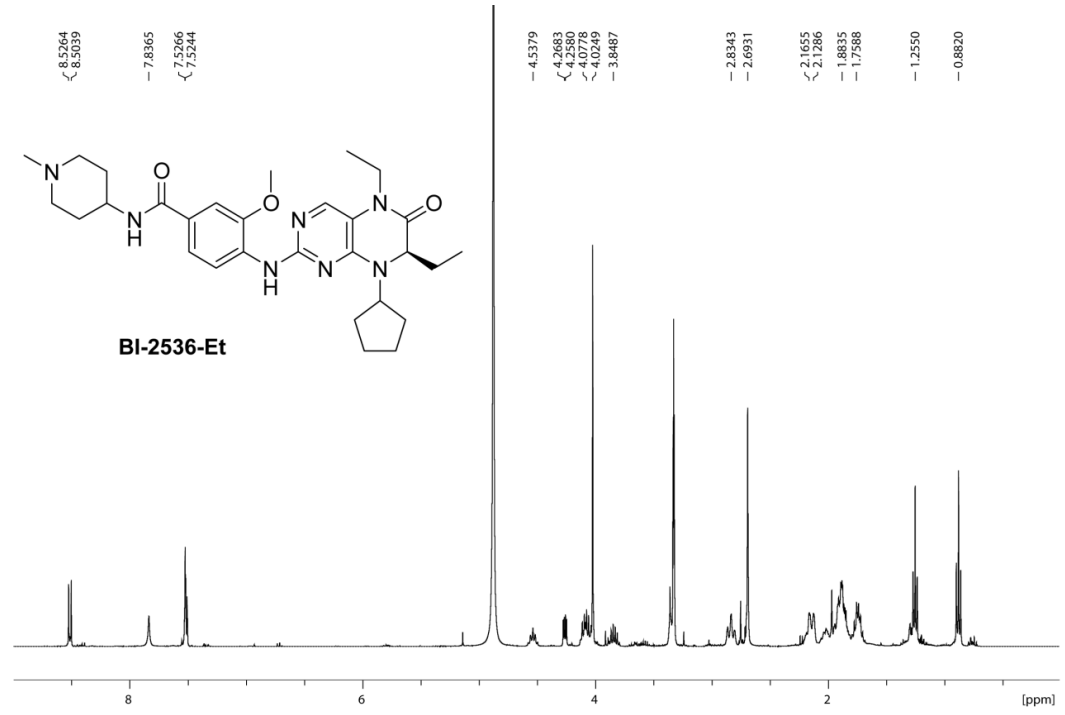
^{1}H NMR of 4-Amino-3-methoxy-N-(1-methylpiperidin-4-yl)benzamide (500 MHz, MeOD)



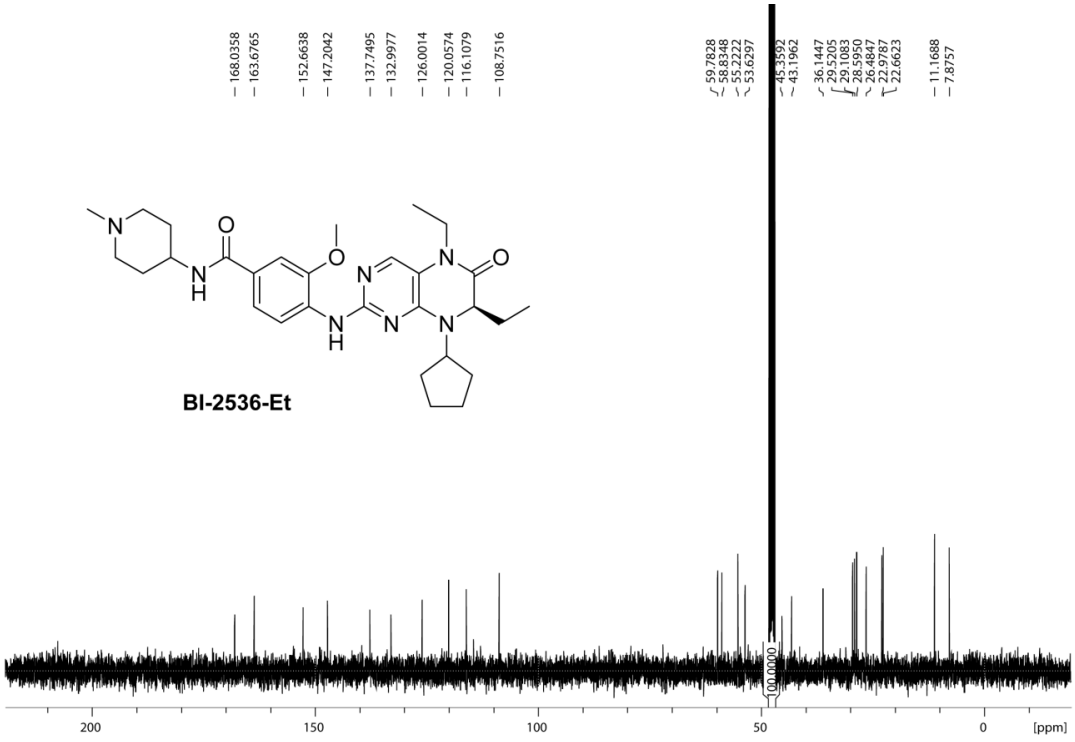
^{13}C NMR of 4-Amino-3-methoxy-N-(1-methylpiperidin-4-yl)benzamide (125 MHz, MeOD)



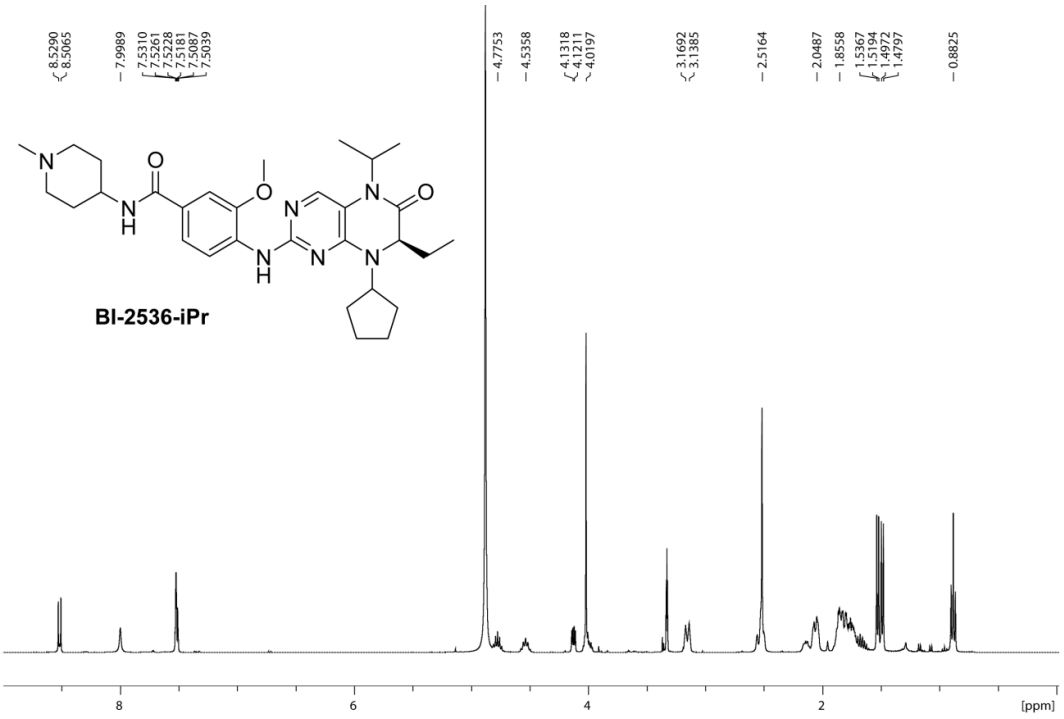
^{1}H NMR of **BI-2536-Et** (500 MHz, MeOD)



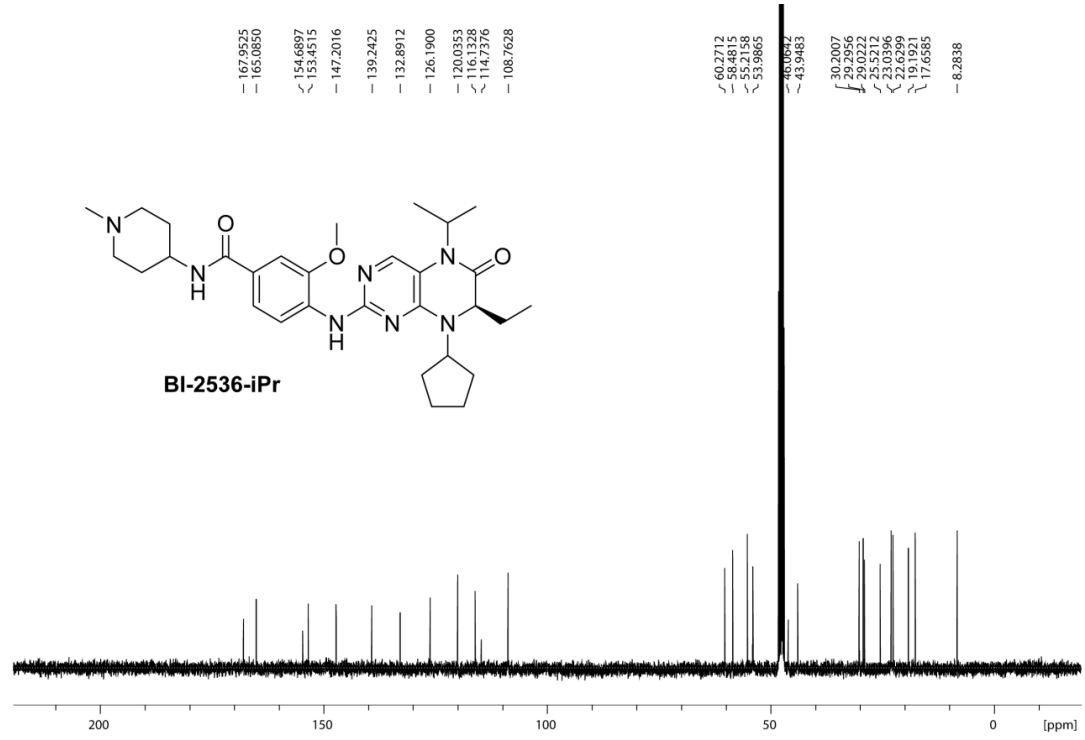
^{13}C NMR of **BI-2536-Et** (125 MHz, MeOD)



^{1}H NMR of **BI-2536-iPr** (500 MHz, MeOD)



^{13}C NMR of **BI-2536-iPr** (125 MHz, MeOD)



Chapter 6. Oxygen Sensing with Perfluorocarbon-Loaded Ultraporous Mesostructured Silica Nanoparticles

Partially reproduced with permission from "Oxygen Sensing with Perfluorcarbon-Loaded Ultraporous Mesostructured Silica Nanoparticles," A. L. Lee*, C. T. Gee*, B. P. Weegman, A. A. Einstein, A. R. Juelfs, H. L. Ring, K. R. Hurley, E. M. Egger, G. Swindlehurst, M. Garwood, W. C. K. Pomerantz, C. L. Haynes, *ACS Nano*, **2017**, *11*, 5623-5632.

Note: This work will also be published in full in the dissertation of Amani Lee.

Motivation: This project was a three-way collaborative project between the Pomerantz, Haynes, and Garwood labs that began as a passing conversation in the NMR lab. It quickly became an opportunity to explore new uses for fluorine magnetic resonance in addressing biological questions. Building on my expertise in fluorine NMR, rather than ligand binding, our goal was to use fluorine NMR and MRI to both quantify the amount of perfluorinated material that was loaded into mesoporous silica nanoparticles and to use these loaded particles for oximetry experiments. My primary contributions to this project were in the quantification of perfluorocarbons in the PERFUMNS and NMR relaxation measurements. As such, portions of the original manuscript pertaining to the synthesis and characterization of the nanoparticles will be described in Appendix 1 so that this chapter focuses on areas of the project to which I directly contributed.

6.1 **Introduction**

Oxygen homeostasis is vital for regulating biological function, including metabolism and disease progression.[268-272] In most cases, hypoxic tumors are resistant to both radiation and many traditional chemotherapy treatments; researchers are developing combinatorial therapies as well as agents activated by tumor hypoxia to treat cancer.[273-275] Additionally, pathogenic infections are highly sensitive to the oxygen level.[271-272] Better measurements of oxygen levels in these tissue microenvironments are needed to optimize

oxygen dependent therapies. Given the significance of tissue oxygenation, many researchers have sought to develop efficient and sensitive technologies for oximetry measurements.[276] However, current oximeters tend to be invasive, consume oxygen, or are limited to superficial measurements.[271, 277] The combination of fluorine-19 magnetic resonance imaging (^{19}F MRI) and biocompatible nanomaterials reported herein has the potential to be an effective platform for in situ oxygen detection.

Noninvasive oximetry methods that measure oxygen levels are currently being explored for clinical applications. Magnetic resonance is a method that uses electromagnetic radiation to probe systems without a limitation in tissue penetration. Techniques such as electron paramagnetic resonance (EPR) spectroscopy, nuclear magnetic resonance (NMR) spectroscopy, magnetic resonance imaging (MRI), and magnetic resonance spectroscopy (MRS) are able to detect the levels of oxygen under strong magnetic fields.[278-279] EPR is a spectroscopic technique that can detect species with unpaired electrons.[280] Clinically approved EPR spectrometers are being developed and optimized with careful considerations of the microwave radiation needed for excitation.[281-284] Alternatively, MRI and NMR detect signals from NMR-active nuclei. For in vivo oximetry, ^{19}F MRI and fluorine-19 MRS (^{19}F MRS), using highly fluorinated materials, are particularly attractive approaches and have already been used for enhanced imaging,[285-286] cell tracking,[287] and probing the microenvironments of tissue.[288]

The ^{19}F nucleus is 100% naturally abundant, has a high gyromagnetic ratio (γ = 40.05 MHz/T), and is highly responsive to changes in its chemical environment. The use of ^{19}F nuclei is also advantageous due to the lack of endogenous fluorine and thus background noise.[289-290] In ^{19}F MRI, partially or perfluorinated compounds (i.e., perfluorocarbons, PFCs), commonly used as contrast agents, have the capacity to dissolve high amounts of gases (e.g., O_2, N_2, and CO_2).[291-293] Upon dissolution into the PFC, the presence of O_2 changes the spin−lattice relaxation time constant (T_1) of the ^{19}F nuclei, based on dipole−dipole interactions between the ^{19}F nucleus and the O_2 unpaired electrons, which is detected by NMR in a concentration-dependent manner. PFCs used for oxygen sensing or delivery have been established in the literature, including perfluoro-15-crown-5-ether (PFCE), hexafluorobenzene (HFB), perfluorodecalin (PFD),

perfluorooctyl bromide (PFOB), perfluoro(tert-butylcyclohexane) (PFTBCH), and perfluorotributylamine (PFTBA).[294-299]

The current state-of-the-art method for dispersing PFCs in aqueous media, as must be done for in vivo imaging, is the emulsification of PFCs with surfactants.[298, 300-301] However, these emulsions can become unstable and lead to coalescence and Ostwald ripening, effectively restricting homogeneous dispersion of the probe.[302-303] To increase the colloidal stability and homogeneous distribution, recent studies have encapsulated PFCs into biocompatible mesoporous silica nanoparticles (MSNs). The advantages of MSNs are their controllable pore size and surface chemistry; their internal space is suitable for loading cargo of interest, and their surfaces can be modified with biological targeting moieties. Chin et al. report a silica nanocapsule filled with PFD for ultrasound measurements.[304] Thermogravimetric analysis and ^{19}F NMR showed that 78 wt % of particles consisted of PFD, whereas the remaining mass consisted of the silica shell. While the silica nanocapsule stabilized the PFD, in comparison to the emulsion, the colloids destabilized completely after 24 h. Nakamura et al. reported a PFC core−shell MSN called mFLAME filled with PFCE for drug delivery.[305] The mFLAME particles were modified with different surface ligands which allowed for dual-mode detection using ^{19}F MRI and fluorescence microscopy. The authors were able to obtain in vitro ^{19}F MR images from 2 mM PFCE concentration at 11.7 T with an acquisition time of approximately 30 min (1820 s). They also obtained in vivo measurements in a mouse model with an acquisition time as fast as 250 s. However, the core−shell structure limits both the amount and versatility of loadable PFC per nanoparticle.

Here, we are evaluating the loading capacity of ultraporous mesostructured silica nanoparticles (UMNs) as an alternative material for ^{19}F MRI. The insensitivity of MRI requires high concentrations of fluorine to overcome the detection limit. It is estimated that 30 mM of ^{19}F may be needed for clinical application (≤3 T).[286] Thus, the loading capacity of MSNs and the PFC fluorine density need careful consideration. The current MSNs have pore volumes much lower than 2 $cm^3\ g^{-1}$, limiting the volume of PFCs that can be loaded. We recently reported on the synthesis of UMNs, spherical MSNs that feature pore volumes up to 4.5 $cm^3\ g^{-1}$.[306] The surfaces of UMNs are readily modified

with polyethylene glycol (PEG) and chlorotrimethylsilane (TMSCl) to improve their colloidal stability and biocompatibility as biologically inert vehicles. Given their comparatively high pore volume, stability under a variety of conditions, and prior precedent with silica nanoparticles, these particles were selected as an ideal candidate for loading with perfluorocarbons.

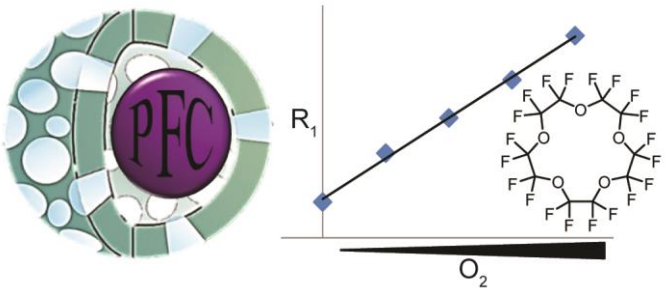


Figure 6.1 Developing a perfluorocarbon (PFC)-loaded ultraporous mesostructured silica nanoparticles.
(Left) as an oximetry tool by monitoring relaxation with sample perflurocarbon perfluoro-15-crown-5-ether (right)

In this study, we developed perfluorocarbon-loaded ultraporous mesostructured silica nanoparticles (PERFUMNs) as ^{19}F-MRI-detectable oxygen-sensing probes. We assessed the loading efficiency of PERFUMNs loaded with three different PFCs, with ^{19}F NMR spectroscopy, and compared the performance to existing MSNs. As our initial attempt to study our fluorinated contrast agents in a biologically relevant context, we evaluated the PERFUMNs' stability in whole blood over several days. Finally, we demonstrated the sensitivity of PERFUMNs to oxygen with ^{19}F NMR relaxometry.

6.2 **Results and Discussion**

Dual-modified UMNs were synthesized and characterized based partially on previously reported procedures.[306] Further details on the synthesis, characterization, and loading of perfluorocarbons can be found in Appendix 1

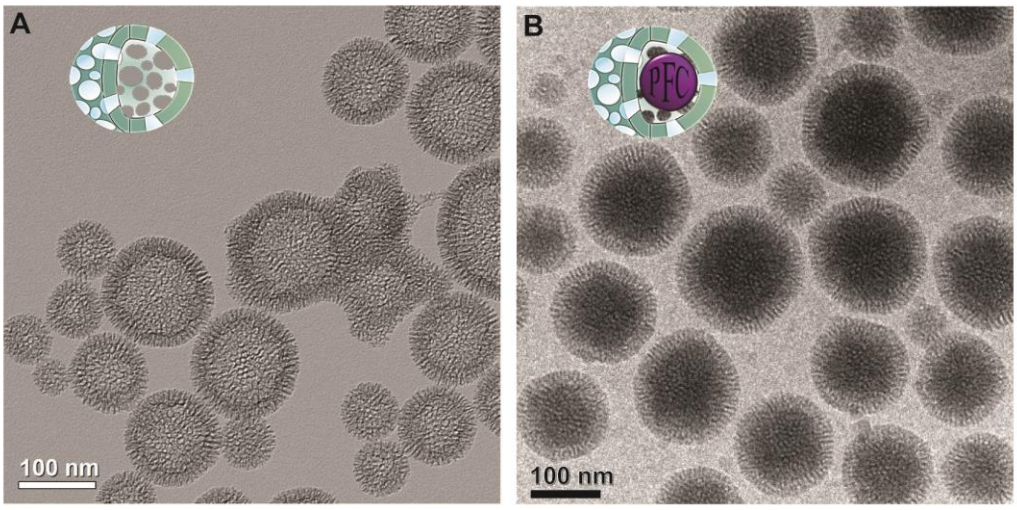


Figure 6.2 Cryo-TEM images of UMNS (A) before and (B) after loading with perfluorocarbons.

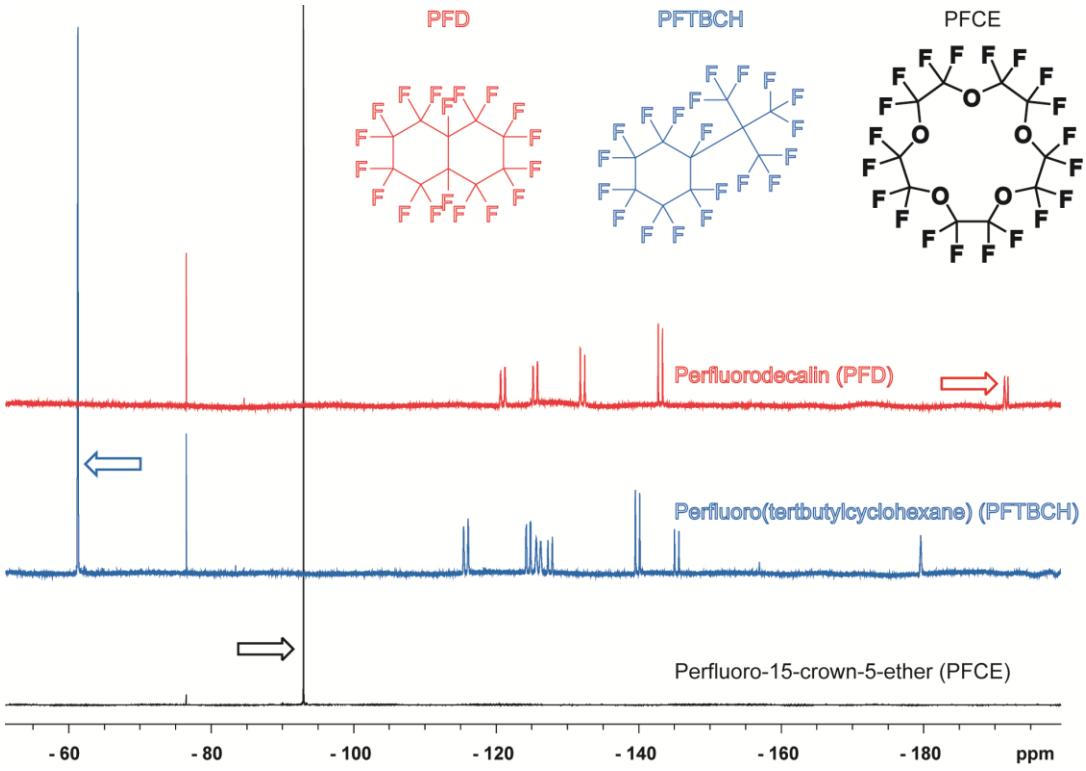


Figure 6.3 Stacked ^{19}F NMR full spectra of three different PERFUMNs showing all resonances for each perfluorocarbon.
All spectra are referenced to TFA at -76.5 ppm. All spectra were obtained at 470 MHz on a Bruker Avance III 500 MHz instrument equipped with a 5 mm Prodigy TCI Cryoprobe. Arrows indicate resonances corresponding to T_1 values in Table 6.1.

6.2.1 ^{19}F NMR Quantification of PERFUMN Loading.

The loading capacity of the PERFUMNs was calculated through quantitative NMR. For maximizing ^{19}F MRI signal sensitivity, it is necessary to take into account not only the amount of perfluorocarbon loaded into the nanoparticles but also how much signal each sample provides per magnetically equivalent fluorine atom. For the materials under study here, PFCE contains 20 magnetically equivalent fluorine; PFTBCH contains 9 equivalent fluorine, whereas PFD contains 4 equivalent fluorines.[307-308] PERFUMNs were loaded in triplicate, and spectra for each PERFUMN were obtained (Figure 6.3 and Figure 6.4). The three different PERFUMNs, loaded with the PFCs above, were compared by the amount of fluorine per mass of nanoparticles (Figure 6.4, inset). Importantly, confinement in our UMNs does not affect the line widths of the observed fluorine resonances compared with the neat PFC resonances, supporting a highly mobile liquid phase within the UMNs (Figure 6.5). The PFCE−PERFUMNs had significantly higher loading than PFD−PERFUMNs ($p < 0.05$) and loaded five times more than PFCE-filled MSNs ($p < 0.05$). The variation of the fluorine loading values of PFTBCH−PERFUMNs made them indistinguishable from the other three samples. Of note, the groups had different sample sizes, 3:6:9:3, from left to right, respectively. However, the one-way ANOVA has some power to account for unequal variances.

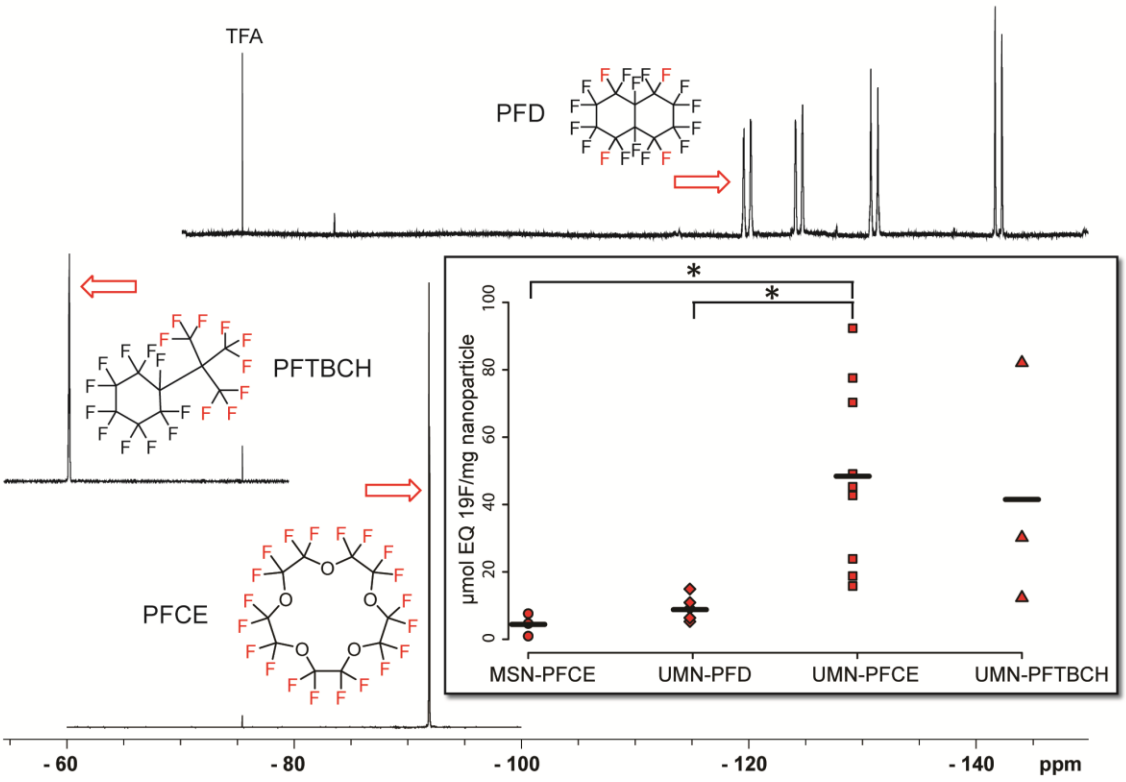


Figure 6.4 Stacked ^{19}F NMR spectra of three different PERFUMNs.
Equivalent fluorines used for quantification are highlighted in the structure and designated with an arrow in each spectrum. All spectra are referenced to trifluoroacetic acid at −76.5 ppm. Inset: PFC loading comparison with MSN particles (circle) and the three type of PERFUMNs (PFD = diamond, PFCE = square, PFTBCH = triangle). The strip chart reveals the spread of the data around the mean (solid line). The total numbers are 3:6:9:3 going from left to right. The perfluorochemical loading for PFCE−PERFUMNs is significantly higher than that for PFD−PERFUMNs and PFCE-loaded MSNs (*$p < 0.05$). Loading of PFTBCH−PERFUMNs is indistinguishable from the other samples.

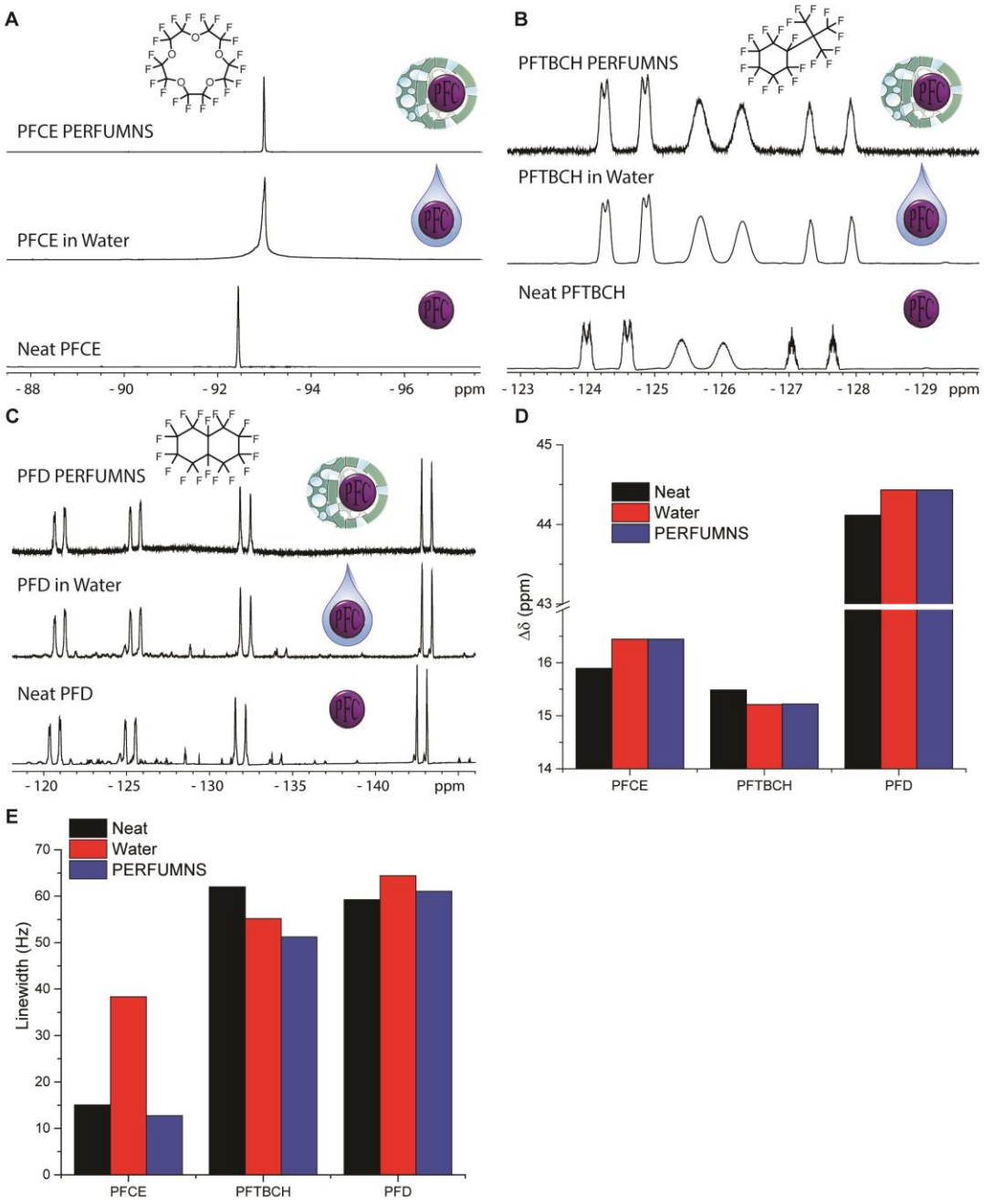


Figure 6.5 Evaluation of PFCs with and without nanoparticles.
(A-C) Corresponding representative ^{19}F NMR spectra in a neat solution (bottom), suspended in water (middle), and loaded in nanoparticles (top). (D) Chemical shift differences of select resonances for each perfluorocarbon relative to trifluoroacetic acid (-76.55 ppm) showing modest changes between neat PFCs and PFCs in water or nanoparticles. (E) Linewidths of select resonances for each perfluorocarbon showing modest differences between neat PFCs and PFCs in water or nanoparticles. It should also be noted that because PFCs are not soluble in water, lineshapes and consequently linewidths can be affected due to poor shimming quality.

PFC-loaded nanoparticles can address the stability challenge of emulsions of PFCs; however, the concentrations of PFCs are lower and particle-size-dependent. In the context of a PFCE emulsion, Ruiz-Cabello et al. reported cationic and anionic emulsion particles with DLS diameters of 158 nm (0.119 PDI) and 207 nm (0.09 PDI), respectively.[286] In 8 μL volumes, the emulsion contained 4.8 mM PFCE. This concentration corresponds to 38 nmol PFCE (760 nmol of ^{19}F). In comparison, in 8 μL, our PFCE−PERFUMNs, with a nanoparticle concentration of 1 mg mL^{-1}, would have 19 nmol PFCE (380 nmol ^{19}F). Although not stated in the article, assuming the 4.8 mM corresponds to the anionic particles, then the 65 nm difference between their larger emulsion and our PERFUMNs led to a 2-fold increase in molar concentration of PFCE. Matsushita et al. have core−shell PFC-filled MSNs synthesized around 80 nm emulsions of PFCE.[309] In 0.5 mL, they reported 1.3 mM PFCE; this concentration corresponds to 0.65 μmol PFCE. In the same volume, Ruiz-Cabello et al. would have 30 μmol PFCE, and our PERFUMNs would have 1.21 μmol PFCE. While increasing the size of the nanoparticles may seem desirable to obtain more fluorine per nanoparticle, the change in size may affect the colloidal stability or biodistribution of the UMNs.[310] As stated before, emulsions have to battle against coalescence and Ostwald ripening; the trade-off for PFC concentration is the added stability of the UMNs.

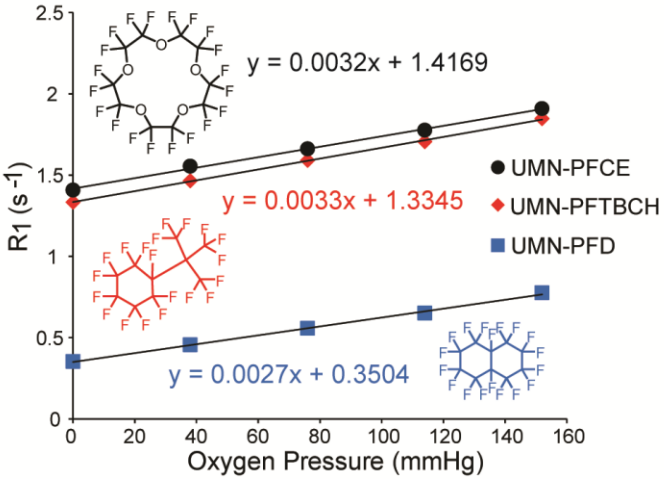


Figure 6.6 Oxygen dependence of R_1 values for different PERFUMNs, demonstrating the linear relationship between R_1 and partial pressure of O_2 in all three cases.
PERFUMNs were measured at 656.8 MHz (16.4 T) at room temperature. Error bars from triplicate measurements are too small to be visible on this graph.

6.2.2 Oxygen Sensitivity of PERFUMNs.

The presence of pores on the surface of PERFUMNs makes the internal liquid PFCs accessible to molecular oxygen outside the nanostructures. As the oxygen dissolves into the liquid PFC, the dipole−dipole interactions between the fluorine nuclei and unpaired electrons of the oxygen reduce the T_1 relaxation times of the nuclei. Traditionally, T_1 values are determined using an inversion− recovery sequence. By measuring the T_1 at different partial pressures of oxygen (pO_2), the relaxation rate constant (R_1), or the inverse of the T_1, can be plotted as a function of pO_2 to give an oxygen calibration curve.[311] Figure 6.6 shows the R_1 as a function of pO_2 at 22 °C, 16.4 T (tuned to 656.8 MHz). As expected, R_1 relaxation rates increase as the concentration of oxygen increases. The slopes and their error were determined with a linear regression fit. PFCE−PERFUMNs had a sensitivity to oxygen with a slope of 322 ± 4 (10^{-5} $mmHg^{-1}$ s^{-1}). PFTBCH−PERFUMNs had a sensitivity of 334 ± 3 (10^{-5} $mmHg^{-1}$ s^{-1}), and PFD−PERFUMNs had the lowest sensitivity of the three PERFUMNs described herein at 273 ± 3 (10^{-5} $mmHg^{-1}$ s^{-1}). To analyze the difference in sensitivities, an analysis of covariance was used to compare the slopes. There was a significant difference between the slopes ($p < 0.05$). A Newman-Keuls multiple comparison test, at a significance level of 0.01, revealed that PFD−PERFUMN sensitivity is significantly different from that of PFTBCH ($p < 0.0001$) and PFCE ($p < 0.0001$), whereas there is no significant difference between the most sensitive PERFUMNs loaded with PFTBCH and PFCE.

In our previous study, we report neat PFCE and PFD have a sensitivity of 223.4 ± 15 (10^{-5} $mmHg^{-1}$ s^{-1}) and 188.0 ± 38 (10^{-5} $mmHg^{-1}$ s^{-1}), respectively, at 16.4 T (37 °C).57 During that study, we also found neat PFCE and PFD have a sensitivity of 280.4 ± 13 (10^{-5} $mmHg^{-1}$ s^{-1}) and 197.3 ± 31 (10^{-5} $mmHg^{-1}$ s^{-1}), respectively, at 16.4 T, 22 °C (not reported). The trend in the intercepts of the PERFUMNs is consistent with the relationship of the inherent T_1 relaxation times between PFCE, PFD, and PFTBCH, where PFD has the longest T_1 relaxation time (Table 6.1). The PERFUMNs have a higher anoxic R_1 compared to that of neat PFC, indicating shorter T_1 relaxation times. The T_1 relaxation times are sensitive to the magnetic field, temperature, and the PFC microenvironment. In this case, the nanoparticles themselves could be affecting the

relaxation of PFCs, likely due to both a decrease in molecular mobility of PFC inside of UMNs and the molecular interactions between the SiO_2 cage and PFC.[312-313] Degradation of the silica nanoparticles may change the response curve of PFC to oxygen. As an initial evaluation of the PERFUMN stability, a preliminary analysis in whole blood is described in the section PERFUMN Oxygen Measurements in Whole Blood.

Table 6.1 ^{19}F-NMR and ^{19}F-MRS Longitudinal Relaxation Times (T_1) of PERFUMNS

System	Temperature	PFCE	PFTBCH	PFD
500 MHz (11.7 T)	300 K	0.775 ± 0.004 s	0.772 ± 0.001 s	1.302 ± 0.006 s
700 MHz (16.4 T)	295 K	0.524 ± 0.001 s	0.541 ± 0.001 s	1.290 ± 0.002 s

^{19}F NMR and ^{19}F MRS measurements were done with 500 MHz (11.7 T) and 700 MHz (16.4 T) systems respectively. Error was reported from instrument as standard deviations. The relaxation time of TFA was also recorded on the 500 MHz system; TFA had relaxation time of 2.659 ± 0.004 s. Refer to Figure S6 for the corresponding PFC resonance being measured.

Our previous study shows oxygen concentration, as well as temperature, can influence the R_1 relaxation rates of fluorine nuclei.[314] We have also shown this phenomenon with PFCE− PERFUMNs (Figure 6.7). As the temperature increased (22 to 45 °C), the O_2 sensitivity of the PFCE−PERFUMNs dropped from 322 (10^{-5} $mmHg^{-1}$ s^{-1}) to 161 (10^{-5} $mmHg^{-1}$ s^{-1}). The anoxic R_1 (0 mmHg O_2) shows PFCE−PERFUMNs have a sensitivity of −226 (10^{-4} °C s^{-1}). If the error in temperature is ±1 °C, this could change the sensitivity to oxygen by 7 (10^{-5} $mmHg^{-1}$ s^{-1}). The variability in temperature, similar to in vivo conditions, adds to the uncertainty in oxygen quantification in a predictable fashion.[315-317] One way to address this issue is to find the oxygen and temperature dependence of each PFC separately; this work is currently underway for future applications.

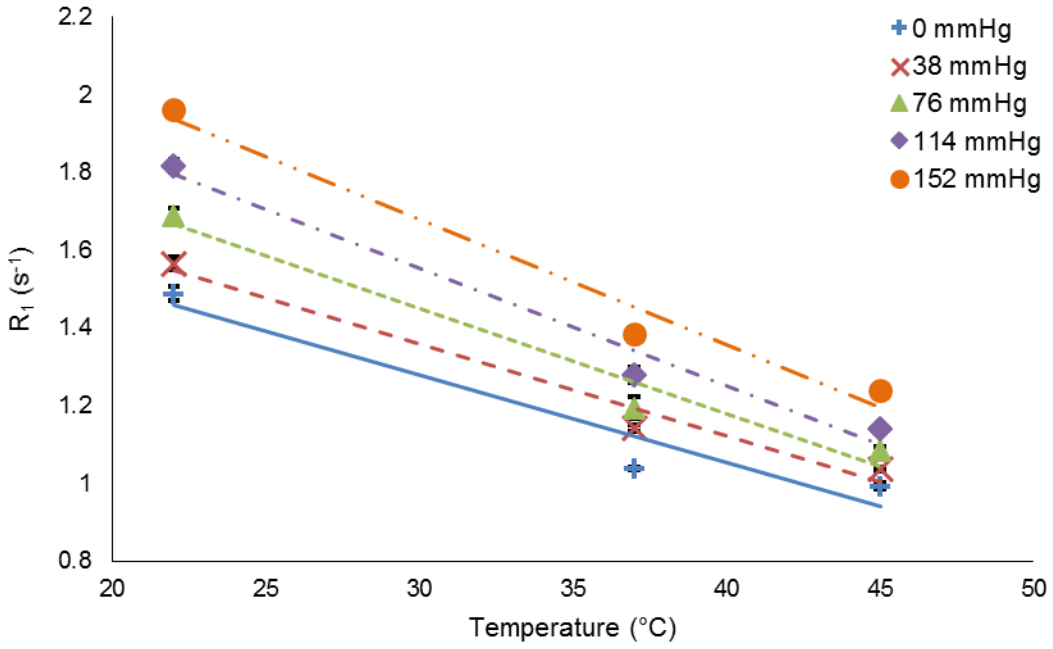


Figure 6.7 Temperature dependence of PFCE-PERFUMNs at 16.4 T.
Oximetry measurements were taken at 22, 37, and 45 °C. Six replicate measurements were taken for each concentration of oxygen: 0 mmHg (+, solid line), 38 mmHg (X, long dash line), 76 mmHg (triangle, small dash line), 114 mmHg (diamond, dot-dash line), and 152 mmHg (circle, dash-dot-dot line). Temperature is shown to have a negative relationship with the PFC relaxation rates (R_1) as the temperature doubles the relaxation rates decrease by 25%. Of note, these PFCE-PERFUMNs are from a different batch than the batch used in Figure 6.6.

In addition to the acquisition of ^{19}F NMR spectra, PERFUMNs were imaged by ^{1}H and ^{19}F MRI at 16.4 T. Comparison of the ^{1}H MRI and ^{19}F MRI images of PFCE−PERFUMNs shows a decrease in the intensity of ^{19}F signal consistent with the 2-fold dilution series (Figure 6.8). The 4 × 4 region of interest (ROI) analysis shows that the signal-to-noise ratio has a linear relationship with the fluorine concentration (Figure 6.9). The linear regression was a good fit (R^2 = 0.95), and the SNR increased by a factor of 2.3 ± 0.3 with change in fluorine concentration. ROI analysis also reveals all clearly visible dilutions have an SNR greater than 10; this is a recommended threshold to overcome at any magnetic field strength.[318] At this magnetic field strength, the lowest concentration (0.025 mg/mL, 1.6 mM equivalent ^{19}F nuclei) of PFCE−PERFUMNs was barely visible,

indicative that concentrations of PERFUMNs ≥ 0.05 mg/mL with greater than 3.1 mM equivalent ^{19}F nuclei are desirable for further oximetry measurements and imaging.

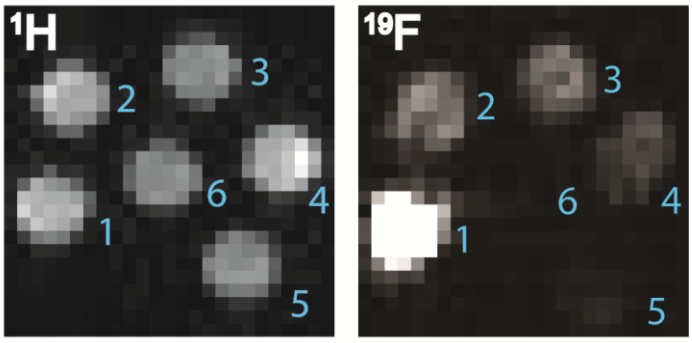


Figure 6.8 MRI phantom images of a PFCE_PERFUMN 2 fold dilution series acquired at 16.4 T and room temperature.
^{1}H and ^{19}F images were acquired in 1.3 s and 27.3 min, respectively. (Left) ^{1}H MRI image. (Right) ^{19}F MRI image. Concentration of equivalent ^{19}F nuclei in 0.8 mL: (1) 24.8 mM (0.4 mg/mL); (2) 12.4 mM (0.2 mg/mL); (3) 6.2 mM (0.1 mg/mL); (4) 3.1 mM (0.05 mg/mL); (5(1.6 mM (0.025 mg/mL); (6) 0 mM (1 mg/mL UMN)

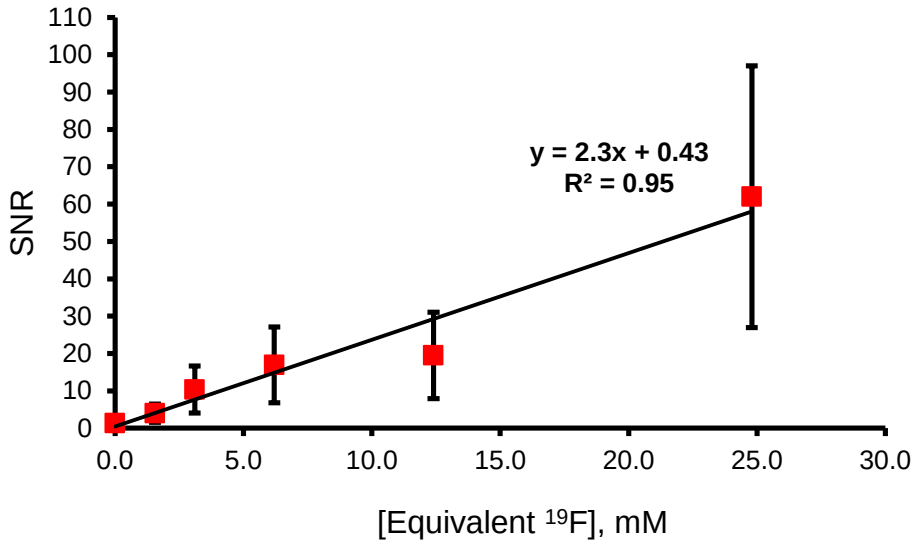


Figure 6.9 Region of interest analysis of 2-fold serial PFCE-PERFUMN dilutions from ^{19}F MR imaging at 16.4 T (656.8 MHz).
The concentrations that were visible had a signal-to-noise ratio greater than or equal to 10. Linear regression of the dilutions shows that the SNR increases by a factor of 2.3 ± 0.3 (mean ± SD) when the concentration of fluorine increases. Error bars are the standard deviation.

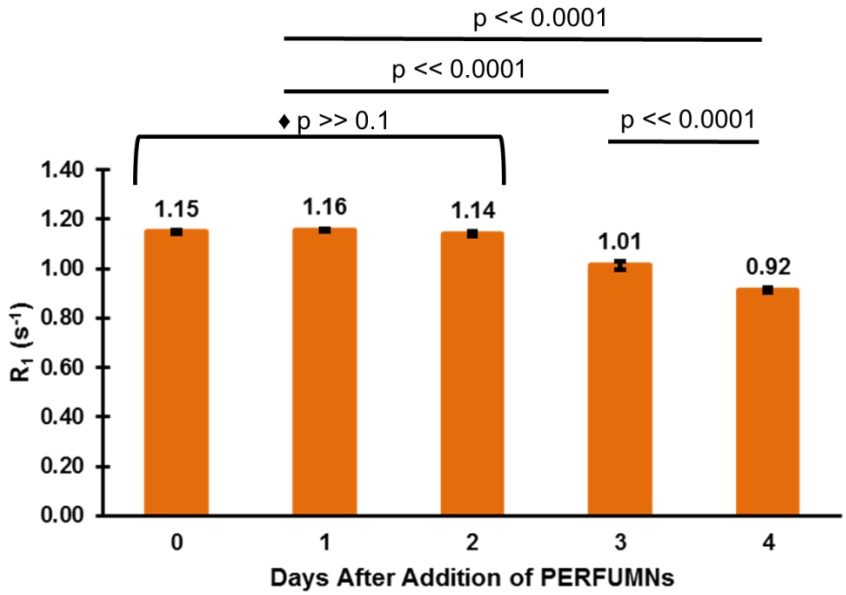


Figure 6.10 Relaxation rates of PFCE-PERFUMNs in whole rabbit blood. Triplicate samples of rabbit blood were treated with PFCE-PERFUMNs, on day 0, and R_1 values were obtained by inversion recovery for 5 days at 470 MHz (11.7 T), 21% O_2, 300 K. The average R_1 values are displayed above the bar graphs. The error bars are standard deviations between triplicate samples. Days 0-2 showed no significant change in relaxation; however, days 3 and 4 are significantly different from days 0-2 and each other. Significance was determined with one-way ANOVA and Tukey multiple comparison test ($\alpha = 0.05$).

6.2.3 PERFUMN Oxygen Measurements in Whole Blood.

Mesoporous silica nanoparticles are known to be stable in biological environments, especially in the hour time frame of experiments previously described.[309, 319-320] To test the feasibility of the PERFUMNs in biologically relevant conditions at long time scales, the oxygen sensitivity of PFCE−PERFUMNs was measured in whole blood. All PFCE−PERFUMN blood measurements were acquired at 470 MHz, 300 K. Triplicate measurements of PFCE−PERFUMNs reveals the R_1 values were constant up to 2 days after incubation in whole blood but started to decrease after 3 days (Figure 6.10). In a second experiment, to measure the sensitivity of PFCE−PERFUMNs to varying oxygen levels, R_1 relaxation rates were measured 2 days after nanoparticles were added to whole blood and compared to PFCE−PERFUMNs in water. Measurements of R_1 at 0, 21, and

100% O_2 reveal a linear dependence but a 1.6- fold decrease in sensitivity to oxygen in whole blood relative to water (Figure 6.11).

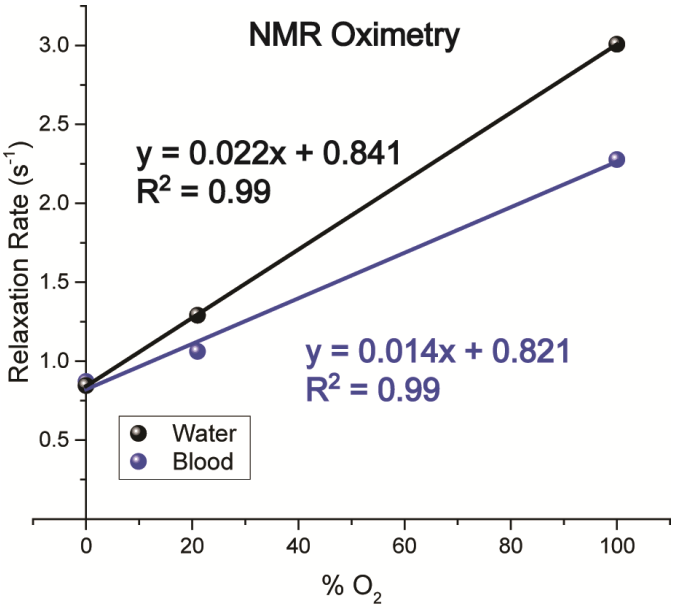


Figure 6.11 Oximetry dependence of R_1 values for PFCE-PERFUMNS in water (top) and in whole rabbit blood (bottom).
Oximetry measurements were taken at 300 K, 470 MHz (11.7 T) at 0%, 21% and 100% O_2. The measurements in blood were taken two days after blood was treated with PERFUMNS

Of note, the rabbit blood was not treated with anticoagulation or anti-inflammatory agents. Potential blood degradation was apparent based on severe discoloration after week long agitation or introduction of anoxic conditions (Figure 6.12). The potential degradation of the blood sample could have affected the relaxation of the PERFUMNs by decreasing the total blood oxygen content. However, for at least up to 2 days, it is clear that oxygen measurements are possible in biological samples. The stability of PERFUMNs and the effects on oxygen sensitivity in biological environments will be the focus of future experiments.

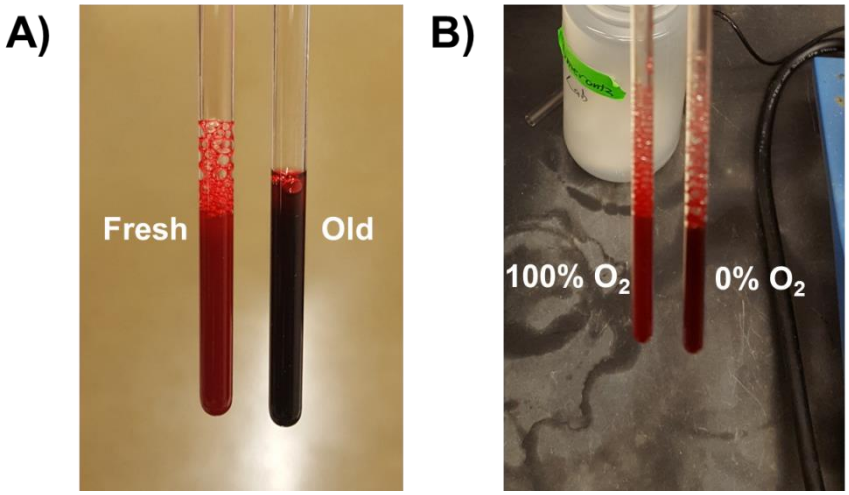


Figure 6.12 Visible change in the integrity of PFCE-PERFUMN blood samples. (A) Difference between blood samples after a week of agitation and freshly drawn blood samples. (B) Discoloration between PFCE-PERFUMN blood samples aerated with 100% O_2 and 0% O_2 (or 100% N_2) for 30 minutes. The discoloration observed is qualitatively similar to blood samples treated with 10% bleach for 30 minutes, indicating that the blood was degrading or becoming unstable.

6.3 **Conclusions**

In summary, we have reported the use of mesoporous silica nanoparticles loaded with perfluorocarbon small molecules for application to oximetry. These particles feature a comparatively high loading capacity, making it possible to encapsulate a large amount of perfluorocarbons (five times greater than MSNs). Additionally, the postsynthesis loading of PFCs allows for use of a broad range of PFC liquids with a single platform nanoparticle carrier. The ability to analytically measure the oxygenation level in biological environments, which was shown feasible in whole blood, is important in the design or improvement of biomedical devices and clinical therapies (e.g., radiation therapy). PERFUMNs are ideal oximetry agents due to their noninvasive NMR-based detection, biologically inert and compatible material, and sensitivity to changes in oxygen levels in microscopic environments. Further exploration of this technology's application in a wide range of settings is underway.

6.4 **Materials and Methods**

6.4.1 Materials.

Perfluoro-15-crown-5-ether was purchased from Exfluor Research Corporation (Round Rock, TX, USA). Perfluorodecalin was purchased from Fluoromed L.P. (Round Rock, TX, USA). Perfluoro(tert-butylcyclohexane) was purchased from Synquest Laboratories, Inc. (Alachua, FL, USA). All chemicals were used without further purification.

6.4.2 Data Analysis.

All statistical analyses were performed with Prism version 6.7 (GraphPad Software, San Diego, CA, USA). Quantitative TEM measurements were performed with ImageJ (NIH, USA). Measurements were made by drawing a line segment across the scale bar and setting the pixel/nm scale. The diameter of the particles was measured by using the oval draw tool to fit an ellipse diameter. Unless stated otherwise, all quantitative data are expressed as the mean ± standard deviation.

6.4.3 TEM Images

Electron micrographs of the UMNs and PERFUMNs were obtained via TEM (FEI Tecnai 12, Houston, TX, USA) and cryo-TEM (Tecnai G2 Spirit Biotwin, Houston, TX, USA), respectively. In TEM, 3 μL of UMNs was dried on Formvarcoated carbon grids overnight. For cryo-TEM, 3 μL of PERFUMNs was placed on lacey carbon grids, which were immediately dried with filter paper for 5 s at 100% humidity. The samples were vitrified by submerging grids into liquid ethane. Electron micrographs were obtained at 120 kV with an emission current of 4 mA.

6.4.4 ^{19}F NMR Quantification.

All NMR spectra were obtained at 470 MHz on a Bruker Avance III 500 MHz instrument equipped with a 5 mm Prodigy TCI cryoprobe. Samples were prepared with 5% D2O and 52 μM trifluoroacetic acid (−76.5 ppm) as a reference and calibration standard to calculate loading values. Loaded nanoparticle solutions were vortexed for 0.5 to 1 min immediately prior to mixing with D2O and TFA. NMR tubes were vortexed for

30 s within 10 min prior to acquisition of NMR spectra. Parameters for ^{19}F NMR experiments included an 8.5 s delay time and a 2 s acquisition time to allow for complete relaxation of magnetization following a full 90° pulse. PFD spectral width and offset were 80 ppm and −110 ppm, respectively. PFCE spectral width and offset were 30 ppm and −85 ppm respectively. PFTBCH, spectral width and offset were 25 ppm and −67 ppm, respectively. Spectra were acquired at 296 K in 64 scans. PFC loading was calculated by comparing relative integrations of select resonances from the PFC to TFA. T_1 relaxation times were measured using the inversion recovery method at 470 MHz and 300 K for determining optimal delay times to allow for quantitative measurement of each fluorine resonance. These measurements were performed under standard atmospheric condition.

6.4.5 ^{19}F MRS Oxygen Sensitivity.

^{19}F spectra and images were acquired with a 16.4 T (tuned to 656.8 MHz). MR system (Agilent Technologies, Santa Clara, CA, USA) using a custom-built single-loop surface coil (2.5 cm diameter) tuned to 656.8 MHz. Oximetry and thermometry methods were previously described for neat compounds.[314] Briefly, the selected PERFUMN was suspended in water and pipetted into a 2 mL vial. The vial was centered in the coil; the coil was tuned and inserted into the isocenter of the magnet. PFCs have been known to saturate with a gas after 15 min.[292-293] The sample was then bubbled and equilibrated with a gas of known oxygen concentration for 30 min [as previously verified with a calibrated 250 μm fluorescence-quenching fiber optic [FO] probe (Ocean Optics, Dunedin, FL, USA)] while the temperature was held constant. Temperature was monitored with an additional FO probe and was maintained using a forced-air heater and water circulator. The magnet was shimmed, and R_1 ($1/T_1$) was then measured with a standard inversion−recovery pulse sequence with customized delay times optimized to each sample's R_1.[321-322] PFCE benefits from processing only a single resonance, but for PFD and PFTBCH (which have multiple resonances), the Ith[323] and major (described above) resonances, respectively, were interrogated. R_1 measurements were made for five different oxygen concentrations (0, 38, 76, 114, and 152 mmHg) at three different temperatures (22, 37, and 45 °C). Each condition was measured in replicate (N = 2−6)

with a repetition time of at least $5/R_1$. R_1 was assumed linear with both oxygen concentration and temperature as previously described.[297, 314, 316]

6.4.6 MRI Phantom Images.

Samples of UMN and PERFUMNs loaded with PFCE were transferred to 6 mm culture tubes. These tubes vertically centered within a 2.5 cm diameter surface coil for imaging. The central tube was the unloaded UMN at 1 mg/mL. A completely empty tube was included in the outer ring of samples between the highest and lowest PERFUMN concentration (0.03 mg UMN/mL and 0.08 mM PFCE). The ^{19}F and ^{1}H images were both acquired using a gradient echo pulse sequence with a resolution of 0.78 × 0.78 mm at 16.4 T. The ^{19}F image was acquired with the following settings: slice thickness = 10 mm; repetition time (TR) = 400 ms; echo time (TE) = 2.0 ms; flip angle = 90°; 64 averages, acquisition time (AQ) = 3.23 ms; receiver bandwidth (sw) = 19841 Hz; data matrix = 64 × 64; and FOV = 50 × 50 mm. The measurement took 27.3 min. The ^{1}H image was acquired with the following settings: slice thickness = 5 mm; TR = 40 ms; TE = 0.78 ms; flip angle = 2°; 1 average; AQ = 0.15 ms; sw = 208333 Hz; data matrix = 32 × 32; and FOV = 25 × 25 mm. The measurement took 1.3 s.

6.4.7 Rabbit Blood Oximetry.

Blood was drawn from New Zealand White rabbits purchased from Bakkom Rabbitry. In the first week, triplicate samples were made by placing 2.5 mL of rabbit blood in a round-bottom polystyrene culture tube with a snap cap. Three tubes with 8 mL of PFCE−PERFUMNs were centrifuged for 12 min at 22 °C and 66 000g. A soft pellet was formed, and 7 mL of supernatant was removed from the tubes. The remaining 1 mL of PFCE− PERFUMNs was vortexed for 20 s, until well mixed, and then added to the tube of rabbit blood. The treated rabbit blood was agitated on a benchtop rocking table at room temperature. Each day, 475 μL of blood was added to NMR tubes with 25 μL of D2O. All ^{19}F NMR measurements were obtained at 470 MHz on a Bruker Avance III 500 MHz instrument at 300 K. T_1 relaxation times were measured using a standard inversion recovery method with time points ranging from 0.010 to 8 s, an acquisition time of 2 s, and delay time of 6 s. In the second week, 2.5 mL of fresh rabbit blood was placed in the

culture tubes. Without centrifugation, 1 mL of PFCE−PERFUMNs from stock was added to the rabbit blood. The blood was placed on the benchtop rocking table at room temperature. Two days after the PERFUMNs were added, 475 μL of treated blood was placed into three J-Young NMR tubes, with 25 μL of D_2O each. One tube was aerated with 100% O_2, and the other was aerated with 100% N_2. Both tubes were aerated for 30 min, and the headspace above the blood was subsequently filled with the corresponding gas to maintain an oxygen or nitrogen-rich environment. The third tube was left open to atmospheric air. The T_1 relaxation times of these samples were measured as described above.

Chapter 7. Quantifying Protein Concentrations Using Smartphone Colorimetry: A New Method for an Established Test

Motivation: As a side project and due to my interest in teaching and education, I undertook the project of developing a new lab experiment utilizing smartphone colorimetry, as established by the Penn lab, in the context of protein colorimetric assays, such as the Bradford and biuret assays. In making these experiments possible with the use of a few easily accessible reagents and smartphone cameras, more students would have access to and the opportunity to learn about proteins and their vital role in biology.

7.1 **Introduction**

Proteins are large biomacromolecules that play necessary roles in nearly all biological processes, including cell growth, metabolism, and differentiation. Consequently, significant research efforts are dedicated to studying the biological function of proteins and designing synthetic molecules to modulate the function of proteins in the context of disease. High school students encounter proteins in both the life sciences as well as in their chemistry courses (e.g., Next Generation Science Standards (NGSS)[324]).

One example of an interesting protein is the bromodomain containing protein 4, (Brd4). When functioning properly, Brd4 is a necessary protein for regulating gene transcription and cell growth, but when functioning improperly, Brd4 has been linked to cancer,[189] a topic covered in life sciences classes and a reality that has impacted many people personally.

Accurate quantification of protein concentration is crucial to studying the roles proteins play in biological processes. This is particularly important in biochemical experiments (e.g., circular dichroism and enzymatic assays) for quantitative characterization of protein structure and function. There are several different methods for protein quantification. Some methods are colorimetric assays that involve the formation of a colored complex. Quantitation is accomplished by determining visible light absorption and using Beer's Law to convert absorbance to concentration by comparison to a standard calibration curve. Two such colorimetric assays are the Bradford and Biuret assays.[325] Using known standards within concentration ranges that obey Beer's law, one can generate a linear calibration curve that can be used to estimate a protein concentration of an unknown sample. These methods are particularly useful when the protein concentration in solution cannot be determined by other techniques., particularly in cases when the protein extinction coefficient is unknown or is too low for direct determination of the concentration based on the protein solution's absorbance at 280 nm.[326]

Colorimetric assays are typically performed using a UV-Vis Spectrophotometer. In the presence of varying concentrations of protein, absorbance values at 595 nm and 540 nm, are monitored in the Bradford and Biuret assays respectively. However, these specialized instruments are expensive and may not be available in laboratories such as teaching laboratories in high schools. To mitigate this challenge, recent reports have demonstrated the efficacy of handheld cameras for quantitative colorimetry, making colorimetric experiments accessible to a wider range of students and researchers.[327] Smartphone colorimetry has been used with kinetics experiments with crystal violet,[328] quantification of gold nanoparticles present in dietary supplements[329] and a variety of other experiments.[330-334] These examples involve a single color, with a colorless blank. The use of camera phone colorimetry with a color change has yet to be explored. Here, we extend quantitative colorimetry to protein assays that involve color changes and its application to quantify protein concentration (Figure 7.1).

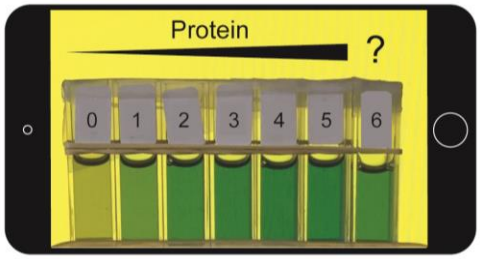


Figure 7.1. Concept figure depicting smartphone colorimetry.

Bradford assays are commonly employed in laboratory experiments for introductory biochemistry courses. The protein-responsive molecule in the Bradford solution, Coomassie Blue G-250, is an environmentally sensitive triphenylmethane dye compound that exists in varying charge states depending on the surrounding solution conditions. The hydrophobic interactions provided by the presence of a protein stabilizes the anionic form of the dye, resulting in color change from a brownish green, with a maximum absorption at 470 nm (cationic form), to a blue color, with a maximum absorption at 595 nm (anionic form) (Figure 7.2).

While the Bradford assay is a dye-based assay, the Biuret assay involves the formation of a Cu(II) complex with the amide backbone present in peptides and proteins. This coordination complex results in a color change from blue (free Cu^{2+}), with a maximum absorption at 650 nm, to violet (HN---Cu^{2+}---NH), with a maximum absorption at 540 nm (Figure 7.2). The Bicinchoninic Acid (BCA) and Lowry assays are examples of other copper-based assays that are commercially available.

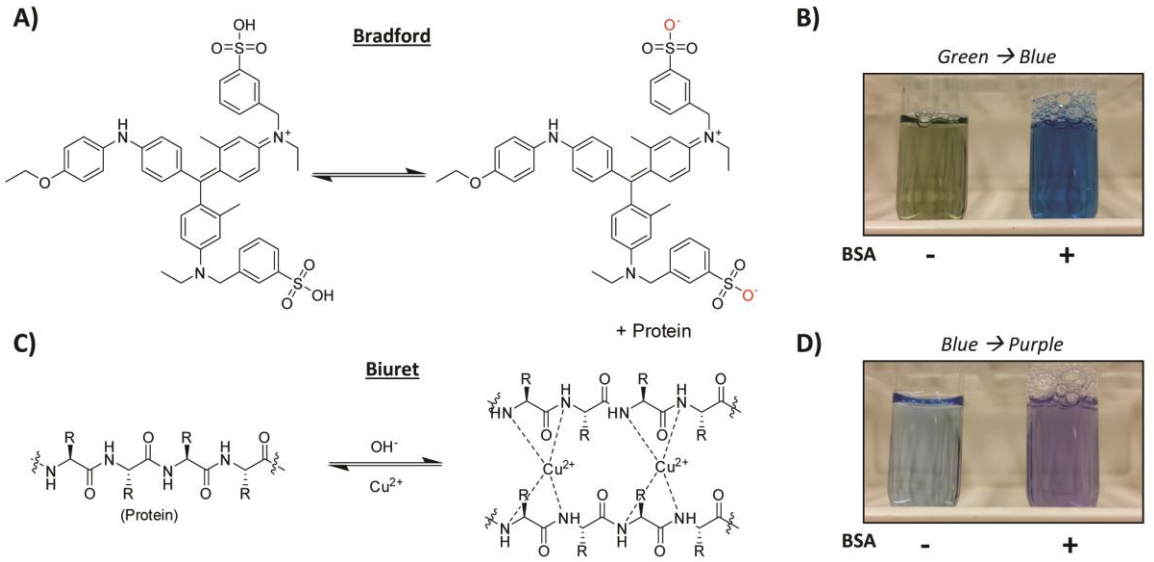


Figure 7.2. **Chemical scheme and sample images for protein colorimetric assays.** A) Chemical scheme for Coomassie Blue G-250 which becomes fully deprotonated in the presence of a protein, yielding a blue color. B) Sample image of Bradford reagent plus water (left) and Bradford reagent plus 0.04 mg/mL Bovine Serum Albumin (BSA) (right). C) Chemical scheme for complexation of cupric ions in the presence of peptides and proteins form a complex with the amide backbone, resulting in a color change from blue to violet. D) Sample image of Biuret reagent plus water (left) and Biuret reagent plus 0.75 mg/mL BSA (right)

Herein we report the first application of smartphone colorimetry for protein quantification using two well-known protein assays that involve changes in color as a function of protein concentration. The ubiquity of the Bradford assay as a standard laboratory experiment in biochemistry and the cheap and accessible nature of the Biuret assay combined with the clear color change that occurs upon complexing with protein made these two assays ideal for use with smartphone colorimetry.

7.2 **Materials and Methods**

7.2.1 Preparation of Chemicals

The Bradford reagent was prepared with Coomassie Blue G-250 (Sigma), methanol (Sigma), phosphoric acid (Macron), and water according to the following protocol: 1) Dissolve 50 mg Coomassie Blue G-250 in 50 mL methanol. 2) Add 100 mL 85% (w/v) phosphoric acid. 3) Add this mixture slowly to 850 mL of water. 4) Store away from light

at 4° C. Because Flinn is a common vendor for high school laboratories, the Biuret reagent was prepared using the Flinn Scientific specifications (0.2% (w/v) $CuSO_4$, 30% (w/v) NaOH) for this reagent. This solution can be made by mixing 100 mg $CuSO_4 \cdot 5H_2O$ (dissolved in 10 mL of water) and 40 mL of 37.5% (w/v) NaOH in water. It is important to note that Biuret from Flinn contains 30% sodium hydroxide while many other vendors and recipes employ only 3% (w/v) sodium hydroxide with the addition of 0.6% (w/v) potassium sodium tartrate.[335] Both solutions will provide adequate results, but the tartrate-containing solution performs significantly better than the solution without it at protein concentrations above 0.5 mg/mL because it prevents protein precipitation. If tartrate is not available, citrate may also be used as a substitute to prevent turbidity (See Supplemental Information). While the Bradford reagent is more common in biochemistry labs for its greater sensitivity to a wider protein concentration range, Biuret is roughly 10% of the cost of Bradford, making it more accessible for a high school or a low-budget laboratory. Bovine serum albumin (BSA) for the calibration curve was purchased from RPI and used as is. Brd4 was expressed from recombinant DNA using standard protein expression protocols in BL21 (DE3) *E. coli* cells.[98]

7.2.2 Preparation of Calibration and "Unknown" Solutions

Standard Bradford or Biuret solutions for the calibration curve were prepared by diluting 0.10 mg/mL BSA (for Bradford assay) or 3 -10 mg/mL BSA (for Biuret assay) with 1.5 mL of Bradford or Biuret solution and water. To prepare the calibration standards, 1.5 mL of the Bradford or Biuret solutions; 0.1, 0.2, 0.3, 0.4, and 0.5 mL of appropriate protein stock solution, and enough water to yield a final volume of 2.0 mL were added to individual vials. The blank sample was prepared by mixing 1.5 mL of the Bradford or Biuret and 0.5 mL of water. Experiments can all be performed with regular tap water. As noted in the sample student guide (Supporting Information), these low volumes of protein solution can be delivered into the individual square plastic cuvettes (1 cm pathlength) by counting drops of the solution and weighing the sample using an balance after each addition to account for the actual amounts added. This approach works well since the densities of each solution and water are all nearly identical. Further, if students do not have access to a balance, counting drops and converting volume to mass

can also provide adequate data, as noted in the results section. Absorbance measurements were performed 40 minutes after mixing for the Bradford and 20 minutes after mixing for the Biuret samples in order to ensure full development of the color change. The "unknown" protein concentration sample was prepared in a similar manner, with the estimated concentration falling within the linear range of the colorimetric agents. The "unknown" protein concentration in the sample Bradford assays was 0.014 mg/mL Brd4 as determined by UV absorption (ε_{280}: 26930 M^{-1} cm^{-1}) and 1.25 mg/mL BSA by mass in the sample Biuret assay. Because the change in color is a time-dependent process, the equilibration time is important for the color to fully develop. Insufficient equilibration time will often lead to poorer results.

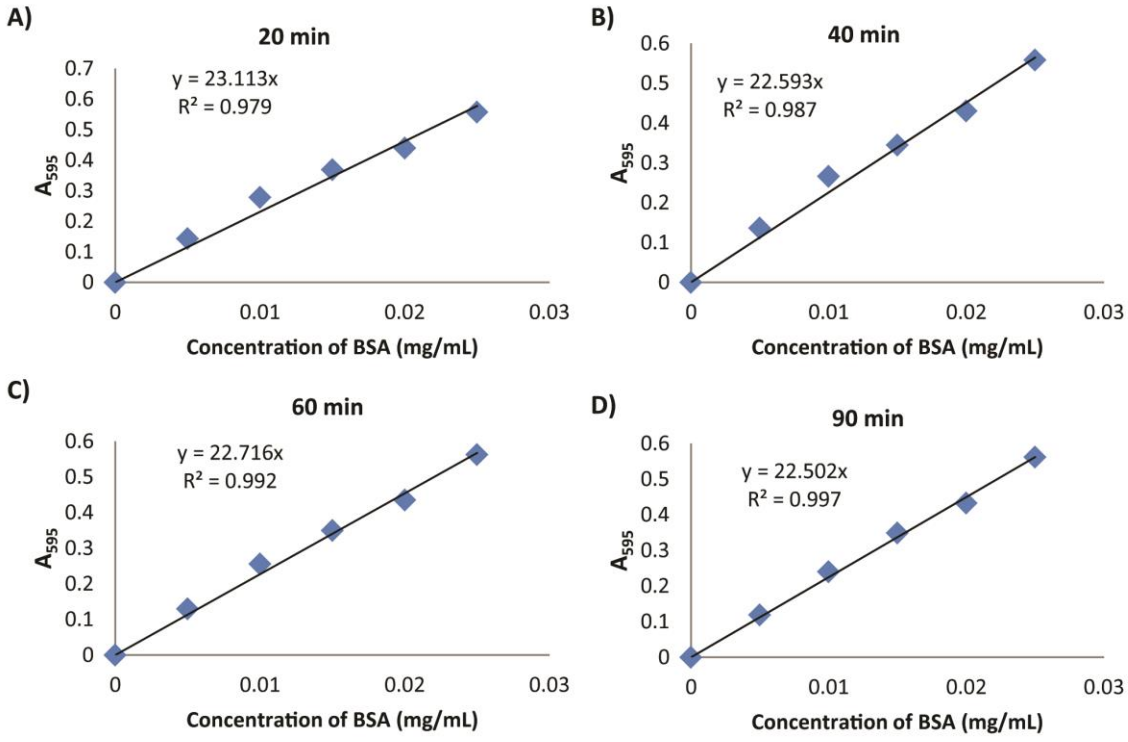


Figure 7.3. Bradford assay data (via UV-Vis) at selected time points demonstrating the time dependent effect on the results.
The samples were measured after approximately A) 20 min, B) 40 min, C) 60 min, and D) 90 min with a gradual increase in R^2 as the data became more linear.

7.2.3 Instrumentation and Setup

Cuvettes containing the standard, blank, and "unknown" samples were analyzed using both a standard UV-Vis Spectrophotometer and smartphone colorimetry. For absorption measurements, samples were analyzed at 595 nm for Bradford and 540 nm for Biuret using a Beckman Coulter DU 720 UV-Vis Spectrophotometer.

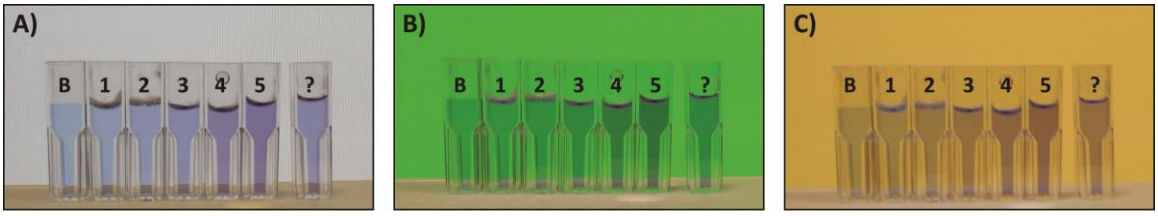


Figure 7.4. **Photographs of calibration samples plus an unknown sample for smartphone colorimetry.**
A) Biuret standards for calibration and "unknown" sample in front of a white background. B) Biuret standards for calibration and "unknown" sample in front of a green (129, 255, 0) background. C) Biuret standards for calibration and "unknown" sample in front of a yellow-orange (255, 207, 0) background.

7.2.4 Acquiring Images

For smartphone colorimetry, cuvettes were lined up in front of a computer monitor, and images were acquired using the back facing lens of varying smartphones (Samsung Galaxy S5, iPhone 6). Cuvettes were lined up in front of a computer monitor with varying background colors matching the wavelengths measured for absorption in addition to a white background (Figure 7.4). For Bradford samples, an RGB code of (255, 207, 0) was used, while (129, 255, 0) was used for the Biuret. These values can be obtained using an online wavelength to RGB conversion tool (e.g. Academo's Wavelength to Colour Relationship[336]).

7.2.5 Data Analysis and Comparison

Absorption data from the spectrophotometer were plotted using spreadsheet software and fitted to a linear trendline. The generated equation was used to estimate the protein concentration. Absorbance measurements from the smartphone images were obtained by analyzing RGB (red, green blue) values from images of each sample (intensity range of

0-255). A variety of software, freeware, and web applications (e.g., Adobe Photoshop, ImageJ, Color Code Picker)[327] as well as smartphone applications (e.g., Colorometer for iPhone and ColorMeter Free for Android)[333] is available for this type of analysis and provide similar results. Download links can be found in the supporting information. RGB values can be obtained by analyzing single points or averaging multiple points over a specified area. Single point and area measurements yield nearly identical results, as noted in a previous report.[327]

Absorbance (A) is calculated from RGB values using the following formula:

$$A = -\log(\frac{I_n}{I_{blank}}) \quad (1)$$

where I_n corresponds to the R, G, or B value of each sample and I_{blank} corresponds to the R, G, or B value for the blank. For the Bradford assay, the best results were obtained using the R values while for the Biuret assay, the best results were obtained using the G values. For alternative colorimetric experiments or assays, it would be best to follow whichever channel (R, G, or B) provides the greatest dynamic range. These "absorbance" values were plotted as a function of concentration and fit to a linear trendline to estimate protein concentration of the "unknown" sample.

7.3 **Hazards**

The Bradford solution is acidic while the Biuret solution is basic. Both are corrosive and should be handled with care. Both solutions should be neutralized appropriately prior to disposing down the drain with excess water. Alternatively, solutions can be disposed of in separate waste containers as appropriate. Protective gloves and eyewear (preferably splash proof goggles) in addition to standard lab-appropriate attire (long pants and closed-toed shoes) should be worn at all times while handling Bradford or Biuret reagents.

7.4 **Results**

Samples from both the Bradford and Biuret assays were evaluated by a UV-Vis spectrophotometer and by smartphone colorimetry to benchmark a newer technique against an established technique. Calibration curves generated R^2 values greater than 0.9,

and using the equation generated from fitting a linear trendline, the concentration of protein in the "unknown" sample was calculated (Figure 7.5).

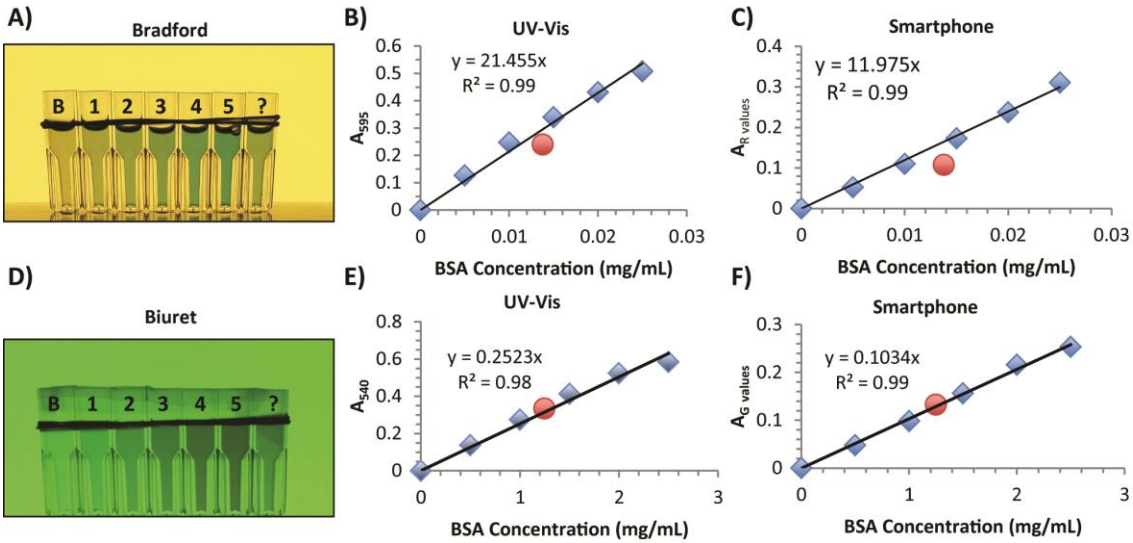


Figure 7.5. Comparison of colorimetric assays by UV-Vis spectrophotometry and smartphone colorimetry.
A) Photograph of Bradford assay cuvettes in front of 595 nm (RGB: 255, 207, 0) colored background. B) Graph of absorbance (via spectrophotometry) versus protein concentration. C) Graph of absorbance (via smartphone colorimetry following R values) versus protein concentration D) Photograph of Biuret assay cuvettes in front of 540 nm (RGB: 129,255,0) colored background. E) Graph of absorbance (via spectrophotometry) versus protein concentration F) Graph of absorbance (via smartphone colorimetry following G values) versus protein concentration In all plots, Blue diamonds refer to protein calibration and the red circle indicates the "unknown" sample at its expected concentration.

7.5 **Discussion**

The use of a smartphone camera is effective for quantitative colorimetry measurements. It can be used for the appearance of color, as previously reported[327-332] and also for a color change, as described here. The obtained data were adequate for correctly estimating a given protein concentration using Beer's Law and a standard calibration curve.

Smartphone colorimetry data were comparable to data collected using the spectrophotometer (Figure 7.5). In the example of the Bradford assay above (Figure 7.5 A-C), the "unknown" concentration was 0.014 mg/mL, and the UV-Vis absorption

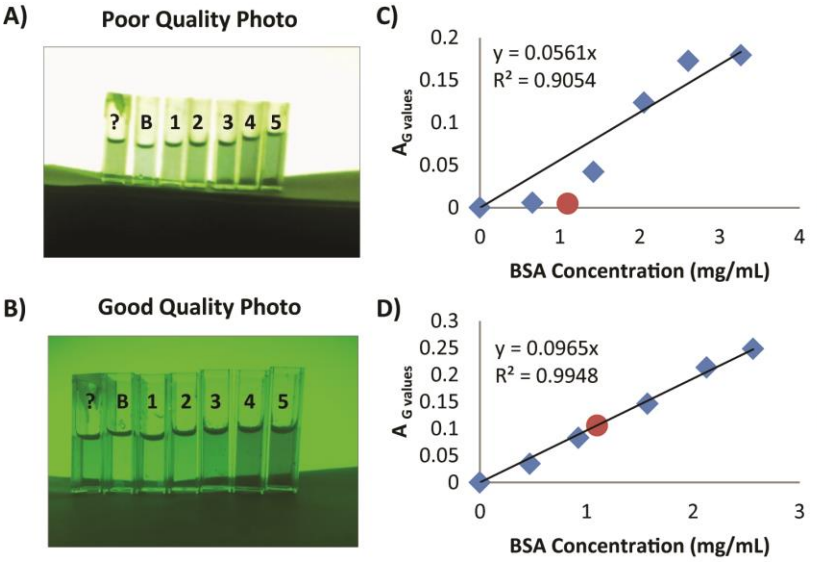
measurements yielded an estimated concentration of 0.011 mg/mL while a concentration of 0.0090 mg/mL was estimated via camera phone colorimetry. In the example of the Biuret assay (using the tartrate-containing solution) (Figure 7.5. D-F), the "unknown" concentration was 1.25 mg/mL, and the UV-Vis absorption measurements yielded an estimated concentration of 1.32 mg/mL while a concentration of 1.28 mg/mL was estimated via camera phone colorimetry. Images acquired in front of backgrounds displaying the color matching the maximum absorption wavelength of the color agent yielded more accurate results as compared to images of the samples arranged in front of a white display. However, if a backlit background is not available, a plain white background is suitable for collecting quantitative measurements. Kuntzleman and Jacobson also reported good success using colored construction paper in lieu of a colored backlit screen for similar types of measurements.[333]

The utility of this method lies not only in its efficacy for determining protein concentration via the measurement of a color change but also in its robustness to a range of factors, including lighting conditions and picture quality, which can vary dramatically from one classroom to another. Given the camera quality found in the average smartphone, this type of protein quantification colorimetry experiment is now more widely accessible to students who may have previously been unable to perform such an experiment.

As a final assessment of the suitability of the experiment for a high school classroom as well as the clarity of the instructions for the experiment, the method was tested with a group of tenth grade high school chemistry students. The lab was conducted with two different class sizes of 25 and 15 students respectively, and run over three 50 minute class periods. The first class period was used to introduce the background for the lab and demonstrate the experimental procedure which can be found in Appendix 2. The second period was used to prepare the solutions and capture the image using smartphones. During the last day, the images were analyzed and the results discussed.

In order to ensure the method is easily reproducible in any high school setting, a few changes were made from the original protocol. For high school implementation we used the Biuret and Albumin as they are available from Flinn Scientific. Also, because

many labs may not have access to protein expression facilities, using an unknown concentration of Albumin, or other commercially available proteins (e.g., lysozyme, or cytochrome C are more practical alternatives than Brd4.) For this study, six groups had access to electronic balances with centigram precision. In order to balance time efficiency and reagents, the students worked in groups of five. The two groups without access to balances were provided with a conversion from drops of solution to mass for each of the solutions.



E)

"Unknown" prepared as 1.1 mg/mL Albumin	
Scale (mg/mL)	Drops (mg/mL)
1.1	1.1
1.1	0.75
1	
0.78	
0.38	
1.7	
Average	Average
1.01	0.925

Figure 7.6. Representative examples of poor quality and good quality student data. A, B) Sample images from students showing a poor quality, washed out image (A) and a clear, usable image (B). C, D) Sample data generated from A and B, respectively, demonstrating the quality of fit for each photo. E) Summary table of student results for measuring the "unknown" sample which was prepared at 1.1 mg/mL Albumin. Each group consisted of five students. A comparison of the student data generated from samples prepared using a balance versus counting drops is also included.

Analysis of the students' data pointed to several areas for further clarification in the protocol (Figure 7.6). At this point, all students had a strong grasp of Beer's law; however, analysis of the results showed of the eight groups performing the experiment, four had unknown concentrations that agreed within ten percent of the unknown protein concentration. The remaining four groups had considerably larger errors. Examples of both preparations of data collection are shown in the supporting information. It should be noted that positive results were independent of sample preparation method. In the four

groups that yielded accurate results and in the four that yielded poor data, three groups prepared samples with a balance and one group counted drops. After reviewing the students' work, it became clear that some groups whose analysis fell outside of 10% of the accepted concentration, had not taken images of suitable quality for analysis. A common error was taking the picture from too far away, leaving the color intensity in the image washed out, as the area of interest was very bright relative to the background. Poor image focus also led to unreliable data. To alleviate these problems for future implementation, providing students with examples of good and poor images have now been added to the instructional guide.

In general, the instructor felt the results were of similar quality to other quantitative methods carried out by high school students at this level. The use of photocolorimetry, Beer's Law, and protein topics, align well with courses introducing content that use the next generation science standards, HS-LS1-1, regarding content on protein function, as well as HS-PS4- 5 which describes how some technological devices use the principles of wave behavior and wave interactions with matter to transmit and capture information and energy.

7.6 **Conclusion**

Smartphones and other portable cameras were successfully used to determine the concentration of a protein sample using assays that yield a color change as a function of protein concentration. This new method for measuring protein concentrations allows for greater accessibility to such colorimetric assays for introductory labs and labs lacking expensive equipment like spectrophotometers. This type of experiment also fits in well with the Science and Engineering Practices from the HS-LS1 section of the NGSS.[324] A detailed step-by-step procedure for the experiment suitable for use in high school or college settings is provided in Appendix 2.

Chapter 8. References

(1) Yan, C.; Higgins, P. J., Drugging the undruggable: Transcription therapy for cancer. *Biochim. Biophys. Acta.* **2013**, *1835*, 76-85.

(2) Stumpf, M. P.; Thorne, T.; de Silva, E.; Stewart, R.; An, H. J.; Lappe, M.; Wiuf, C., Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U.S.A* **2008**, *105*, 6959-6964.

(3) Koehler, A. N., A complex task? Direct modulation of transcription factors with small molecules. *Curr. Opin. Chem. Biol.* **2010**, *14*, 331-340.

(4) Azzarito, V.; Long, K.; Murphy, N. S.; Wilson, A. J., Inhibition of alpha helix mediated protein-protein interactions using designed molecule. *Nat. Chem.* **2013**, *5*, 161-173.

(5) Klein, C. V., L T, Targeting the p53-MDM2 interaction to treat cancer. *Br. J. Cancer* **2004**, *91*, 1415-1419.

(6) Vassilev, L. T., Small-molecule antagonists of p53-MDM2 binding: research tools and potential therapeutics. *Cell Cycle* **2004**, *3*, 419-421.

(7) Vassilev, L. T.; Vu, B. T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C.; Fotouhi, N.; Liu, E. A., In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **2004**, *303*, 844-848.

(8) Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C., Binding of small molecules to an adaptive protein-protein interface. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 1603-1608.

(9) Braisted, A. C.; Oslob, J. D.; Delano, W. L.; Hyde, J.; McDowell, R. S.; Waal, N.; Yu, C.; Arkin, M. R.; Raimundo, B. C., Discovery of a potent small molecule IL-2 inhibitor through fragment assembly. *J. Am. Chem. Soc.* **2003**, *125*, 3714-3715.

(10) Raimundo, B. C.; Oslob, J. D.; Braisted, A. C.; Hyde, J.; McDowell, R. S.; Randal, M.; Waal, N. D.; Wilkinson, J.; Yu, C. H.; Arkin, M. R., Integrating fragment assembly and biophysical methods in the chemical advancement of small-molecule antagonists of IL-2: an approach for inhibiting protein-protein interactions. *J. Med. Chem.* **2004**, *47*, 3111-3130.

(11) Bruncko, M.; Oost, T. K.; Belli, B. A.; Ding, H.; Joseph, M. K.; Kunzer, A.; Martineau, D.; McClellan, W. J.; Mitten, M.; Ng, S. C.; Nimmer, P. M.; Oltersdorf, T.; Park, C. M.; Petros, A. M.; Shoemaker, A. R.; Song, X.; Wang, X.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H.; Elmore, S. W., Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL. *J. Med. Chem.* **2007**, *50*, 641-662.

(12) Hajduk, P. J., SAR by NMR: putting the pieces together. *Mol. Interv.* **2006**, *6*, 266-272.

(13) Oltersdorf, T.; Elmore, S. W.; Shoemaker, A. R.; Armstrong, R. C.; Augeri, D. J.; Belli, B. A.; Bruncko, M.; Deckwerth, T. L.; Dinges, J.; Hajduk, P. J.; Joseph, M. K.; Kitada, S.; Korsmeyer, S. J.; Kunzer, A. R.; Letai, A.; Li, C.; Mitten, M. J.;

Nettesheim, D. G.; Ng, S.; Nimmer, P. M.; O'Connor, J. M.; Oleksijew, A.; Petros, A. M.; Reed, J. C.; Shen, W.; Tahir, S. K.; Thompson, C. B.; Tomaselli, K. J.; Wang, B.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H., An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* **2005**, *435*, 677-681.

(14) Cesa, L. C.; Mapp, A. K.; Gestwicki, J. E., Direct and propagated effects of small molecules on protein-protein interaction networks. *Front. Bioeng. Biotechnol.* **2015**, *3*, 119.

(15) Lee, L. W.; Mapp, A. K., Transcriptional switches: Chemical approaches to gene regulation. *J. Biol. Chem.* **2010**, *285*, 11033-11038.

(16) Thompson, A. D.; Dugan, A.; Gestwicki, J. E.; Mapp, A. K., Fine-tuning multiprotein complexes using small molecules. *ACS Chem. Biol.* **2012**, *7*, 1311-1320.

(17) University, J. H. Online Mendelian Inheritance in Man. http://www.omim.org (accessed 10/5/2013).

(18) Wells, J. A.; McClendon, C. L., Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **2007**, *450*, 1001-1009.

(19) Clackson, T.; Wells, J. A., A hot spot of binding energy in a hormone-receptor interface. *Science* **1995**, *267*, 383-386.

(20) Moreira, I. S.; Martins, J. M.; Ramos, R. M.; Fernandes, P. A.; Ramos, M. J., Understanding the importance of the aromatic amino-acid residues as hot-spots. *Biochim. Biophys. Acta.* **2013**, *1834*, 404-414.

(21) London, N.; Movshovitz-Attias, D.; Schueler-Furman, O., The structural basis of peptide-protein binding strategies. *Structure* **2010**, *18*, 188-199.

(22) Bullock, B. N.; Jochim, A. L.; Arora, P. S., Assessing helical protein interfaces for inhibitor design. *J. Am. Chem. Soc.* **2011**, *133*, 14220-14223.

(23) Watkins, A. M.; Arora, P. S., Anatomy of beta-strands at protein-protein interfaces. *ACS Chem. Biol.* **2014**, *9*, 1747-1754.

(24) Gavenonis, J.; Sheneman, B. A.; Siegert, T. R.; Eshelman, M. R.; Kritzer, J. A., Comprehensive analysis of loops at protein-protein interfaces for macrocycle design. *Nat. Chem. Biol.* **2014**, *10*, 716-722.

(25) Nelson, D. L.; Cox, M. M., *Lehninger Principles of Biochemistry*. 2008.

(26) Darnell, J. E., Jr., Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer* **2002**, *2*, 740-749.

(27) Mapp, A. K.; Ansari, A. Z., A TAD further: Exogenous control of gene activation. *ACS Chem. Biol.* **2007**, *2*, 62-75.

(28) Berg, J. M. T., John L.; Stryer, Lubert, *Biochemistry*. 7th ed.; W. H. Freeman and Company: New York, 2012.

(29) Sakamoto, K. M.; Frank, D. A., CREB in the pathophysiology of cancer: Implications for targeting transcription factors for cancer therapy. *Clin. Cancer. Res.* **2009**, *15*, 2583-2587.

(30) Mayr, B.; Montminy, M., Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat. Rev. Mol. Cell. Biol.* **2001**, *2*, 599-609.

(31) Carlezon Jr, W. A.; Duman, R. S.; Nestler, E. J., The many faces of CREB. *Trends Neurosci.* **2005**, *28*, 436-445.

(32) Benito, E.; Barco, A., CREB's control of intrinsic and synaptic plasticity: implications for CREB-dependent memory models. *Trends Neurosci.* **2010**, *33*, 230-240.
(33) Blobel, G. A., CREB-binding protein and p300: molecular integrators of hematopoietic transcription. *Blood* **2000**, *95*, 745-755.
(34) Lonze, B. E.; Ginty, D. D., Function and regulation of CREB family transcription factors in the nervous system. *Neuron* **2002**, *35*, 605-623.
(35) Shankar, D. B.; Cheng, J. C.; Sakamoto, K. M., Role of cyclic AMP response element binding protein in human leukemias. *Cancer* **2005**, *104*, 1819-1824.
(36) Best, J. L.; Amezcua, C. A.; Mayr, B.; Flechner, L.; Murawsky, C. M.; Emerson, B.; Zor, T.; Gardner, K. H.; Montminy, M., Identification of small-molecule antagonists that inhibit an activator: coactivator interaction. *Proc. Natl. Acad. Sci. U.S.A* **2004**, *101*, 17622-17627.
(37) Wen, A. Y.; Sakamoto, K. M.; Miller, L. S., The role of the transcription factor CREB in immune function. *J. Immunol.* **2010**, *185*, 6413-6419.
(38) Baron, V.; Adamson, E. D.; Calogero, A.; Ragona, G.; Mercola, D., The transcription factor Egr1 is a direct regulator of multiple tumor suppressors including TGF beta 1, PTEN, p53, and fibronectin. *Cancer Gene Ther.* **2006**, *13*, 115-124.
(39) Campbell, K. M.; Lumb, K. J., Structurally distinct modes of recognition of the KIX domain of CBP by Jun and CREB. *Biochemistry* **2002**, *41*, 13956-13964.
(40) Radhakrishnan, I.; Pérez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E., Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:coactivator interactions. *Cell* **1997**, *91*, 741-752.
(41) Pomerantz, W. C.; Wang, N.; Lipinski, A. K.; Wang, R.; Cierpicki, T.; Mapp, A. K., Profiling the dynamic interfaces of fluorinated transcription complexes for ligand discovery and characterization. *ACS Chem. Biol.* **2012**, *7*, 1345-1350.
(42) Majmudar, C. Y.; Højfeldt, J. W.; Arevang, C. J.; Pomerantz, W. C.; Gagnon, J. K.; Schultz, P. J.; Cesa, L. C.; Doss, C. H.; Rowe, S. P.; Vásquez, V.; Tamayo-Castillo, G.; Cierpicki, T.; Brooks, C. L.; Sherman, D. H.; Mapp, A. K., Sekikaic acid and lobaric acid target a dynamic interface of the coactivator CBP/p300. *Angew. Chem. Int. Ed.* **2012**, *51*, 11258-11262.
(43) Frangioni, J. V.; LaRiccia, L. M.; Cantley, L. C.; Montminy, M. R., Minimal activators that bind to the KIX domain of p300/CBP identified by phage display screening. *Nat. Biotechnol.* **2000**, *18*, 1080-1085.
(44) Jiang, M.; Li, B. X.; Xie, F.; Delaney, F.; Xiao, X., Design, synthesis, and biological evaluation of conformationally constrained analogues of naphthol AS-E as inhibitors of CREB-mediated gene transcription. *J. Med. Chem.* **2012**, *55*, 4020-4024.
(45) Li, B. X.; Xiao, X., Discovery of a small-molecule inhibitor of the KIX-KID interaction. *ChemBioChem* **2009**, *10*, 2721-2724.
(46) Li, B. X.; Yamanaka, K.; Xiao, X., Structure-activity relationship studies of naphthol AS-E and its derivatives as anticancer agents by inhibiting CREB-mediated gene transcription. *Bioorg. Med. Chem.* **2012**, *20*, 6811-6820.

(47) Wang, N.; Majmudar, C. Y.; Pomerantz, W. C.; Gagnon, J. K.; Sadowsky, J. D.; Meagher, J. L.; Johnson, T. K.; Stuckey, J. A.; Brooks, C. L., 3rd; Wells, J. A.; Mapp, A. K., Ordering a dynamic protein via a small-molecule stabilizer. *J. Am. Chem. Soc.* **2013**, *135*, 3363-3366.

(48) Bates, C. A.; Pomerantz, W. C.; Mapp, A. K., Transcriptional tools: Small molecules for modulating CBP KIX-dependent transcriptional activators. *Biopolymers* **2011**, *95*, 17-23.

(49) Davis, B. J.; Erlanson, D. A., Learning from our mistakes: The 'unknown knowns' in fragment screening. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 2844-2852.

(50) *Fluorescence Polarization: Technical Resource Guide*. 4th ed.; Invitrogen Corporation: Madison, WI, 2006.

(51) Kenakin, T. P., *Pharmacologic analysis of drug-receptor interaction*. Raven: New York, 1993.

(52) Barrett, P. J.; Chen, J.; Cho, M. K.; Kim, J. H.; Lu, Z.; Mathew, S.; Peng, D.; Song, Y.; Van Horn, W. D.; Zhuang, T.; Sonnichsen, F. D.; Sanders, C. R., The quiet renaissance of protein nuclear magnetic resonance. *Biochemistry* **2013**, *52*, 1303-1320.

(53) Claridge, T. D. W., *High-Resolution NMR Techniques in Organic Chemistry*. 2nd ed.; Elsevier: Oxford, 2009; Vol. 27, p 383.

(54) Pavia, D. L. L., Gary M.; Kriz, George S.; Vyvyan, James R., *Introduction to Spectroscopy*. 4th ed.; Brooks/Cole, Cengage Learning: United States, 2009; p 656.

(55) Miller, A. T. J., *Essentials of Chemical Biology: Structure and Dynamics of Biological Macromolecules*. John Wiley & Sons Ltd. : England, 2008; p 573.

(56) Kwan, A. H.; Mobli, M.; Gooley, P. R.; King, G. F.; Mackay, J. P., Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J.* **2011**, *278*, 687-703.

(57) Schanda, P.; Brutscher, B., Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **2005**, *127*, 8014-8015.

(58) Zhu, G.; Yao, X. J., TROSY-based NMR experiments for NMR studies of large biomolecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *52*, 49-68.

(59) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K., Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U.S.A* **1997**, *94*, 12366-12371.

(60) Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wuthrich, K., NMR analysis of a 900K GroEL GroES complex. *Nature* **2002**, *418*, 207-211.

(61) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W., Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531-1534.

(62) Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E., Cross-correlated relaxation enhanced 1H−13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **2003**, *125*, 10420-10428.

(63) Sheppard, D.; Sprangers, R.; Tugarinov, V., Experimental approaches for NMR studies of side-chain dynamics in high-molecular-weight proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *56*, 1-45.
(64) Tugarinov, V.; Kay, L. E., Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J. Am. Chem. Soc.* **2003**, *125*, 13868-13878.
(65) Tugarinov, V.; Kay, L. E., Side chain assignments of Ile δ1 methyl groups in high molecular weight proteins: An application to a 46 ns tumbling molecule. *J. Am. Chem. Soc.* **2003**, *125*, 5701-5706.
(66) Gardner, K. H.; Rosen, M. K.; Kay, L. E., Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* **1997**, *36*, 1389-1401.
(67) Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E., Selective methyl group protonation of perdeuterated proteins. *J. Mol. Biol.* **1996**, *263*, 627-636.
(68) Mueller, G. A.; Choy, W. Y.; Yang, D.; Forman-Kay, J. D.; Venters, R. A.; Kay, L. E., Global folds of proteins with low densities of NOEs using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein. *J. Mol. Biol.* **2000**, *300*, 197-212.
(69) Nicholson, L. K.; Kay, L. E.; Baldisseri, D. M.; Arango, J.; Young, P. E.; Bax, A.; Torchia, D. A., Dynamics of methyl groups in proteins as studied by proton-detected carbon-13 NMR spectroscopy. Application to the leucine residues of staphylococcal nuclease. *Biochemistry* **1992**, *31*, 5253-5263.
(70) Cai, S.; Seu, C.; Kovacs, Z.; Sherry, A. D.; Chen, Y., Sensitivity enhancement of multidimensional NMR experiments by paramagnetic relaxation effects. *J. Am. Chem. Soc.* **2006**, *128*, 13474-13478.
(71) Lambert, J. B.; Gronert, S.; Shurvell, H. F.; Lightner, D., *Organic Structural Spectroscopy*. 2nd ed.; Pearson Prentice Hall: United States, 2011; p 533.
(72) Bruker QCI-F Cryoprobe. http://www.bruker.com (accessed 10/6/2013).
(73) Reich, H. J. Fluorine NMR Data. http://www.chem.wisc.edu/areas/reich/handouts/nmr/f-data.htm (accessed 10/6/2013).
(74) Kitevski-LeBlanc, J. L.; Prosser, R. S., Current applications of 19F NMR to studies of protein structure and dynamics. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, *62*, 1-33.
(75) Urick, A. K.; Calle, L. P.; Espinosa, J. F.; Hu, H.; Pomerantz, W. C., Protein-observed fluorine NMR is a complementary ligand discovery method to 1H CPMG ligand-observed NMR. *ACS Chem. Biol.* **2016**, *11*, 3154-3164.
(76) Li, H.; Frieden, C., Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and 19F NMR. *Proc. Natl. Acad. Sci. U.S.A* **2007**, *104*, 11993-11998.
(77) Gerig, J. T., Fluorine NMR of proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **1994**, *26, Part 4*, 293-370.
(78) Suzuki, Y.; Brender, J. R.; Soper, M. T.; Krishnamoorthy, J.; Zhou, Y.; Ruotolo, B. T.; Kotov, N. A.; Ramamoorthy, A.; Marsh, E. N., Resolution of oligomeric

species during the aggregation of Abeta1-40 using (19)F NMR. *Biochemistry* **2013**, *52*, 1903-1912.

(79) Liu, J. J.; Horst, R.; Katritch, V.; Stevens, R. C.; Wuthrich, K., Biased signaling pathways in beta(2)-adrenergic receptor characterized by F-19-NMR. *Science* **2012**, *335*, 1106-1110.

(80) Chung, K. Y.; Kim, T. H.; Manglik, A.; Alvares, R.; Kobilka, B. K.; Prosser, R. S., Role of detergents in conformational exchange of a G protein-coupled receptor. *J. Biol. Chem.* **2012**, *287*, 36305-36311.

(81) Horst, R.; Liu, J. J.; Stevens, R. C.; Wüthrich, K., β2-adrenergic receptor activation by agonists studied with 19F NMR spectroscopy. *Angew. Chem. Int. Ed.* **2013**, *52*, 10762-10765.

(82) Kim, T. H.; Chung, K. Y.; Manglik, A.; Hansen, A. L.; Dror, R. O.; Mildorf, T. J.; Shaw, D. E.; Kobilka, B. K.; Prosser, R. S., The role of ligands on the equilibria between functional states of a G protein-coupled receptor. *J. Am. Chem. Soc.* **2013**, *135*, 9465-9474.

(83) Klein-Seetharaman, J.; Getmanova, E. V.; Loewen, M. C.; Reeves, P. J.; Khorana, H. G., NMR spectroscopy in studies of light-induced structural changes in mammalian rhodopsin: Applicability of solution 19F NMR. *Proc. Natl. Acad. Sci. U.S.A* **1999**, *96*, 13744-13749.

(84) Loewen, M. C.; Klein-Seetharaman, J.; Getmanova, E. V.; Reeves, P. J.; Schwalbe, H.; Khorana, H. G., Solution 19F nuclear Overhauser effects in structural studies of the cytoplasmic domain of mammalian rhodopsin. *Proc. Natl. Acad. Sci. U.S.A* **2001**, *98*, 4888-4892.

(85) Weinstein, S.; Wallace, B. A.; Blout, E. R.; Morrow, J. S.; Veatch, W., Conformation of gramicidin A channel in phospholipid vesicles: a 13C and 19F nuclear magnetic resonance study. *Proc. Natl. Acad. Sci. U.S.A* **1979**, *76*, 4230-4234.

(86) Weinstein, S.; Durkin, J. T.; Veatch, W. R.; Blout, E. R., Conformation of the gramicidin A channel in phospholipid vesicles: fluorine-19 nuclear magnetic resonance study. *Biochemistry* **1985**, *24*, 4374-4382.

(87) Hagen, D. S.; Weiner, J. H.; Sykes, B. D., Fluorotyrosine M13 coat protein: fluorine-19 nuclear magnetic resonance study of the motional properties of an integral membrane protein in phospholipid vesicles. *Biochemistry* **1978**, *17*, 3860-3866.

(88) Wilson, M. L.; Dahlquist, F. W., Membrane protein conformational change dependent on the hydrophobic environment. *Biochemistry* **1985**, *24*, 1920-1928.

(89) Evanics, F.; Bezsonova, I.; Marsh, J.; Kitevski, J. L.; Forman-Kay, J. D.; Prosser, R. S., Tryptophan solvent exposure in folded and unfolded states of an SH3 domain by 19F and 1H NMR. *Biochemistry* **2006**, *45*, 14120-14128.

(90) Li, C.; Lutz, E. A.; Slade, K. M.; Ruf, R. A.; Wang, G. F.; Pielak, G. J., 19F NMR studies of alpha-synuclein conformation and fibrillation. *Biochemistry* **2009**, *48*, 8578-8584.

(91) Wang, G. F.; Li, C.; Pielak, G. J., 19F NMR studies of alpha-synuclein-membrane interactions. *Prot. Sci.* **2010**, *19*, 1686-1691.

(92) Zigoneanu, I. G.; Pielak, G. J., Interaction of alpha-synuclein and a cell penetrating fusion peptide with higher eukaryotic cell membranes assessed by (1)(9)F NMR. *Mol. Pharm.* **2012**, *9*, 1024-1029.
(93) Hoeltzli, S. D.; Frieden, C., Stopped-flow NMR spectroscopy: Real-time unfolding studies of 6-F-19-tryptophan-labeled Escherichia coli dihydrofolatereductase. *Proc. Natl. Acad. Sci. U.S.A* **1995**, *92*, 9318-9322.
(94) Hoeltzli, S. D.; Frieden, C., Real-time refolding studies of 6-F-19-tryptophan labeled Escherichia coli dihydrofolate reductase using stopped-flow NMR spectroscopy. *Biochemistry* **1996**, *35*, 16843-16851.
(95) Hoeltzli, S. D.; Frieden, C., Refolding of 6-F-19 tryptophan-labeled Escherichia coli dihydrofolate reductase in the presence of ligand: A stopped-flow NMR spectroscopy study. *Biochemistry* **1998**, *37*, 387-398.
(96) Bann, J. G.; Pinkner, J.; Hultgren, S. J.; Frieden, C., Real-time and equilibrium 19F-NMR studies reveal the role of domain–domain interactions in the folding of the chaperone PapD. *Proc. Natl. Acad. Sci. U.S.A* **2002**, *99*, 709-714.
(97) Bann, J. G.; Frieden, C., Folding and domain−domain Interactions of the chaperone PapD measured by 19F NMR. *Biochemistry* **2004**, *43*, 13775-13786.
(98) Mishra, N. K.; Urick, A. K.; Ember, S. W.; Schonbrunn, E.; Pomerantz, W. C., Fluorinated aromatic amino acids are sensitive 19F NMR probes for bromodomain-ligand interactions. *ACS Chem. Biol.* **2014**, *9*, 2755-2760.
(99) Urick, A. K.; Hawk, L. M.; Cassel, M. K.; Mishra, N. K.; Liu, S.; Adhikari, N.; Zhang, W.; Dos Santos, C. O.; Hall, J. L.; Pomerantz, W. C., Dual screening of BPTF and Brd4 using protein-observed fluorine NMR uncovers new bromodomain probe molecules. *ACS Chem. Biol.* **2015**, *10*, 2246-2256.
(100) Seyedsayamdost, M. R.; Reece, S. Y.; Nocera, D. G.; Stubbe, J., Mono-, di-, tri-, and tetra-substituted fluorotyrosines: new probes for enzymes that use tyrosyl radicals in catalysis. *J. Am. Chem. Soc.* **2006**, *128*, 1569-1579.
(101) Kitevski-LeBlanc, J. L.; Al-Abdul-Wahid, M. S.; Prosser, R. S., A mutagenesis-free approach to assignment of 19F NMR resonances in biosynthetically labeled proteins. *J. Am. Chem. Soc.* **2009**, *131*, 2054-2055.
(102) Kitevski-Leblanc, J. L.; Evanics, F.; Scott Prosser, R., Approaches to the assignment of (19)F resonances from 3-fluorophenylalanine labeled calmodulin using solution state NMR. *J. Biomol. NMR* **2010**, *47*, 113-123.
(103) Isley, W. C., 3rd; Urick, A. K.; Pomerantz, W. C.; Cramer, C. J., Prediction of (19)F NMR chemical shifts in labeled proteins: Computational protocol and case study. *Mol Pharm* **2016**, *13*, 2376-2386.
(104) Jiang, Z. X.; Feng, Y.; Yu, Y. B., Fluorinated paramagnetic chelates as potential multi-chromic 19F tracer agents. *Chem. Commun.* **2011**, *47*, 7233-7235.
(105) Chalmers, K. H.; De Luca, E.; Hogg, N. H.; Kenwright, A. M.; Kuprov, I.; Parker, D.; Botta, M.; Wilson, J. I.; Blamire, A. M., Design principles and theory of paramagnetic fluorine-labelled lanthanide complexes as probes for (19)F magnetic resonance: a proof-of-concept study. *Chemistry* **2010**, *16*, 134-148.
(106) Chalmers, K. H.; Botta, M.; Parker, D., Strategies to enhance signal intensity with paramagnetic fluorine-labelled lanthanide complexes as probes for 19F magnetic resonance. *Dalton Trans.* **2011**, *40*, 904-913.

(107) Hawk, L. M. L.; Gee, C. T.; Urick, A. K.; Hu, H.; Pomerantz, W. C. K., Paramagnetic relaxation enhancement for protein-observed 19F NMR as an enabling approach for efficient fragment screening. *RSC Adv.* **2016**, *6*, 95715-95721.

(108) Lee, A. L.; Gee, C. T.; Weegman, B. P.; Einstein, S. A.; Juelfs, A. R.; Ring, H. L.; Hurley, K. R.; Egger, S. M.; Swindlehurst, G.; Garwood, M.; Pomerantz, W. C. K.; Haynes, C. L., Oxygen Sensing with Perfluorocarbon-Loaded Ultraporous Mesostructured Silica Nanoparticles. *ACS Nano* **2017**, *11*, 5623-5632.

(109) Matsumoto, Y.; Jasanoff, A., Metalloprotein-based MRI probes. *FEBS Lett.* **2013**, *587*, 1021-1029.

(110) Sanders, J. K.; Williams, D. H., Shift reagents in NMR spectroscopy. *Nature* **1972**, *240*, 385-390.

(111) Mroue, K. H.; MacKinnon, N.; Xu, J.; Zhu, P.; McNerny, E.; Kohn, D. H.; Morris, M. D.; Ramamoorthy, A., High-resolution structural insights into bone: A solid-state NMR relaxation study utilizing paramagnetic doping. *J. Phys. Chem. B* **2012**, *116*, 11656-11661.

(112) Yamamoto, K.; Caporini, M.; Im, S.; Waskell, L.; Ramamoorthy, A., Shortening spin-lattice relaxation using a copper-chelated lipid at low-temperatures – A magic angle spinning solid-state NMR study on a membrane-bound protein. *J. Magn. Reson.* **2013**, 175-181.

(113) Yamamoto, K.; Xu, J.; Kawulka, K. E.; Vederas, J. C.; Ramamoorthy, A., Use of a copper-chelated lipid speeds up NMR measurements from membrane proteins. *J. Am. Chem. Soc.* **2010**, *132*, 6929-6931.

(114) Yamamoto, K.; Vivekanandan, S.; Ramamoorthy, A., Fast NMR data acquisition from bicelles containing a membrane-associated peptide at natural-abundance. *J. Phys. Chem. B* **2011**, *115*, 12448-12455.

(115) Wickramasinghe, N. P.; Parthasarathy, S.; Jones, C. R.; Bhardwaj, C.; Long, F.; Kotecha, M.; Mehboob, S.; Fung, L. W. M.; Past, J.; Samoson, A.; Ishii, Y., Nanomole-scale protein solid-state NMR by breaking intrinsic H-1 T-1 boundaries. *Nat. Methods* **2009**, *6*, 215-218.

(116) Friebolin, H., *Basic One- and Two-Dimensional NMR Spectroscopy*. 3rd ed.; Wiley-VCH: Federal Republic of Germany, 1998; p 386.

(117) Solomon, I., Relaxation processes in a system of two spins. *Phys. Rev.* **1955**, *99*, 559-566.

(118) Clore, G. M.; Iwahara, J., Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* **2009**, *109*, 4108-4139.

(119) Bloembergen, N.; Morgan, L. O., Proton relaxation times in paramagnetic solutions. Effects of electron spin relaxation. *J. Chem. Phys.* **1961**, *34*, 842-850.

(120) Carr, H. Y.; Purcell, E. M., Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.* **1954**, *94*, 630-638.

(121) Meiboom, S.; Gill, D., Modified spin-echo method for measuring nuclear relaxation Times. *Rev. Sci. Instrum.* **1958**, *29*, 688-691.

(122) Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R., NMR spectroscopy of paramagnetic metalloproteins. *ChemBioChem* **2005**, *6*, 1536-1549.
(123) Bertini, I.; Luchinat, C.; Parigi, G., Magnetic susceptibility in paramagnetic NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *40*, 249-273.
(124) Barry, C. D.; North, A. C.; Glasel, J. A.; Williams, R. J.; Xavier, A. V., Quantitative determination of mononucleotide conformations in solution using lanthanide ion shift and broadenine NMR probes. *Nature* **1971**, *232*, 236-245.
(125) Hull, W. E.; Sykes, B. D., Fluorotyrosine alkaline phosphatase: Internal mobility of individual tyrosines and the role of chemical shift anisotropy as a 19F nuclear spin relaxation mechanism in proteins. *J. Mol. Biol.* **1975**, *98*, 121-153.
(126) Tang, S.; Case, D. A., Calculation of chemical shift anisotropy in proteins. *J. Biomol. NMR* **2011**, *51*, 303-312.
(127) Tjandra, N.; Szabo, A.; Bax, A., Protein backbone dynamics and 15N chemical shift anisotropy from quantitative measurement of relaxation interference effects. *J. Am. Chem. Soc.* **1996**, *118*, 6986-6991.
(128) Sternberg, U.; Klipfel, M.; Grage, S. L.; Witter, R.; Ulrich, A. S., Calculation of fluorine chemical shift tensors for the interpretation of oriented 19F-NMR spectra of gramicidin A in membranes. *Phys. Chem. Chem. Phys.* **2009**, *11*, 7048-7060.
(129) Dürr, U. H. N.; Grage, S. L.; Witter, R.; Ulrich, A. S., Solid state 19F NMR parameters of fluorine-labeled amino acids. Part I: Aromatic substituents. *J. Magn. Reson.* **2008**, *191*, 7-15.
(130) Grage, S. L.; Wang, J.; Cross, T. A.; Ulrich, A. S., Solid-state 19F-NMR analysis of 19F-labeled tryptophan in gramicidin A in oriented membranes. *Biophys. J.* **2002**, *83*, 3336-3350.
(131) Gerig, J., Fluorine NMR. *Biophysics Textbook Online* **2001**, 1-35.
(132) Saito, H.; Ando, I.; Ramamoorthy, A., Chemical shift tensor - the heart of NMR: Insights into biological aspects of proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57*, 181-228.
(133) Begam Elavarasi, S.; Kavita, D., Characterization of the chemical shielding tensor using cross-correlated spin relaxation measurements and quantum chemical calculations. *Chem. Phys. Lett.* **2010**, *489*, 248-253.
(134) Dorai, K.; Kumar, A., Fluorine chemical shift tensors in substituted fluorobenzenes using cross correlations in NMR relaxation. *Chem. Phys. Lett.* **2001**, *335*, 176-182.
(135) Post, J. F.; Cottam, P. F.; Simplaceanu, V.; Ho, C., Fluorine-19 nuclear magnetic resonance study of 5-fluorotryptophan-labeled histidine-binding protein J of Salmonella typhimurium. *J. Mol. Biol.* **1984**, *179*, 729-743.
(136) Williams, S. P.; Haggie, P. M.; Brindle, K. M., 19F NMR measurements of the rotational mobility of proteins in vivo. *Biophys. J.* **1997**, *72*, 490-498.
(137) Sykes, B. D.; Weingarten, H. I.; Schlesinger, M. J., Fluorotyrosine alkaline phosphatase from Escherichia coli: preparation, properties, and fluorine-19 nuclear magnetic resonance spectrum. *Proc. Natl. Acad. Sci. U S A* **1974**, *71*, 469-473.

(138) Shi, P.; Wang, H.; Xi, Z.; Shi, C.; Xiong, Y.; Tian, C., Site-specific (1)(9)F NMR chemical shift and side chain relaxation analysis of a membrane protein labeled with an unnatural amino acid. *Protein Sci.* **2011**, *20*, 224-228.
(139) Lee, H. W.; Sohn, J. H.; Yeh, B. I.; Choi, J. W.; Jung, S. H.; Kim, K. W., F-19 NMR investigation of F-1-ATPase of Escherichia coli using fluorotryptophan labeling. *J. Biochem.* **2000**, *127*, 1053-1056.
(140) Salopek-Sondi, B.; Vaughan, M. D.; Skeels, M. C.; Honek, J. F.; Luck, L. A., (19)F NMR studies of the leucine-isoleucine-valine binding protein: evidence that a closed conformation exists in solution. *J. Biomol. Struct. Dyn.* **2003**, *21*, 235-246.
(141) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug. Deliv. Rev.* **1997**, *23*, 3-25.
(142) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H., A 'Rule of Three' for fragment-based lead discovery? *Drug Discov. Today* **2003**, *8*, 876-877.
(143) Jhoti, H., A new school for screening. *Nat. Biotechnol.* **2005**, *23*, 184-186.
(144) Dalvit, C., NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discov. Today* **2009**, *14*, 1051-1057.
(145) Mishra, K. P.; Ganju, L.; Sairam, M.; Banerjee, P. K.; Sawhney, R. C., A review of high throughput technology for the screening of natural products. *Biomed. Pharmacother.* **2008**, *62*, 94-98.
(146) Mayr, L. M.; Bojanic, D., Novel trends in high-throughput screening. *Curr. Opin. Pharmacol.* **2009**, *9*, 580-588.
(147) Hajduk, P. J.; Greer, J., A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug. Discov.* **2007**, *6*, 211-219.
(148) Carr, R. A. E.; Congreve, M.; Murray, C. W.; Rees, D. C., Fragment-based lead discovery: leads by design. *Drug Discov. Today* **2005**, *10*, 987-992.
(149) Dahlin, J. L.; Nissink, J. W.; Strasser, J. M.; Francis, S.; Higgins, L.; Zhou, H.; Zhang, Z.; Walters, M. A., PAINS in the assay: chemical mechanisms of assay interference and promiscuous enzymatic inhibition observed during a sulfhydryl-scavenging HTS. *J. Med. Chem.* **2015**, *58*, 2091-2113.
(150) Hopkins, A. L.; Keseru, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H., The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov* **2014**, *13*, 105-121.
(151) Gleeson, M. P.; Hersey, A.; Montanari, D.; Overington, J., Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat. Rev. Drug Discov.* **2011**, *10*, 197-208.
(152) Cala, O.; Guilliere, F.; Krimm, I., NMR-based analysis of protein-ligand interactions. *Anal. Bioanal. Chem.* **2014**, *406*, 943-956.
(153) Coles, M.; Heller, M.; Kessler, H., NMR-based screening technologies. *Drug Discovery Today* **2003**, *8*, 803-810.
(154) Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T. L.; Homans, S. W.; Kessler, H.; Luchinat, C.; Meyer, B.; Oschkinat, H.; Peng, J.; Schwalbe, H.; Siegal, G., Perspectives on NMR in drug discovery: a technique comes of age. *Nat. Rev. Drug Discov.* **2008**, *7*, 738-745.

(155) Dalvit, C.; Vulpetti, A., Technical and practical aspects of (19) F NMR-based screening: toward sensitive high-throughput screening with rapid deconvolution. *Magn. Reson. Chem.* **2012**, *50*, 592-597.
(156) Dalvit, C., Ligand- substrate-based 19F NMR screening: Principles and applications to drug discovery. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51*, 243-271.
(157) Zhou, Y.; Wang, J.; Gu, Z.; Wang, S.; Zhu, W.; Acena, J. L.; Soloshonok, V. A.; Izawa, K.; Liu, H., Next generation of fluorine-containing pharmaceuticals, compounds currently in phase II-III clinical trials of major pharmaceutical companies: New structural trends and therapeutic areas. *Chem. Rev.* **2016**, *116*, 422-518.
(158) Lo, M. C.; Aulabaugh, A.; Jin, G.; Cowling, R.; Bard, J.; Malamas, M.; Ellestad, G., Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **2004**, *332*, 153-159.
(159) Niesen, F. H.; Berglund, H.; Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2007**, *2*, 2212-2221.
(160) Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; Salemme, F. R., High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screen.* **2001**, *6*, 429-440.
(161) Molina, D. M.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y.; Nordlund, P., Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **2013**, *341*, 84-87.
(162) Jafari, R.; Almqvist, H.; Axelsson, H.; Ignatushchenko, M.; Lundback, T.; Nordlund, P.; Martinez Molina, D., The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat. Protoc.* **2014**, *9*, 2100-2122.
(163) Makley, L. N.; McMenimen, K. A.; DeVree, B. T.; Goldman, J. W.; McGlasson, B. N.; Rajagopal, P.; Dunyak, B. M.; McQuade, T. J.; Thompson, A. D.; Sunahara, R.; Klevit, R. E.; Andley, U. P.; Gestwicki, J. E., Pharmacological chaperone for α-crystallin partially restores transparency in cataract models. *Science* **2015**, *350*, 674-677.
(164) Giannetti, A. M., From experimental design to validated hits a comprehensive walk-through of fragment lead identification using surface plasmon resonance. *Methods Enzymol.* **2011**, *493*, 169-218.
(165) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **2004**, *3*, 935-949.
(166) Ayoub, A. M.; Hawk, L. M. L.; Herzig, R. J.; Jiang, J.; Wisniewski, A. J.; Gee, C. T.; Zhao, P.; Zhu, J. Y.; Berndt, N.; Offei-Addo, N. K.; Scott, T. G.; Qi, J.; Bradner, J. E.; Ward, T. R.; Schonbrunn, E.; Georg, G. I.; Pomerantz, W. C. K., BET bromodomain inhibitors with one-step synthesis discovered from virtual screen. *J. Med. Chem.* **2017**, *60*, 4805-4817.
(167) Souers, A. J.; Leverson, J. D.; Boghaert, E. R.; Ackler, S. L.; Catron, N. D.; Chen, J.; Dayton, B. D.; Ding, H.; Enschede, S. H.; Fairbrother, W. J.; Huang, D. C.;

Hymowitz, S. G.; Jin, S.; Khaw, S. L.; Kovar, P. J.; Lam, L. T.; Lee, J.; Maecker, H. L.; Marsh, K. C.; Mason, K. D.; Mitten, M. J.; Nimmer, P. M.; Oleksijew, A.; Park, C. H.; Park, C. M.; Phillips, D. C.; Roberts, A. W.; Sampath, D.; Seymour, J. F.; Smith, M. L.; Sullivan, G. M.; Tahir, S. K.; Tse, C.; Wendt, M. D.; Xiao, Y.; Xue, J. C.; Zhang, H.; Humerickhouse, R. A.; Rosenberg, S. H.; Elmore, S. W., ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **2013**, *19*, 202-208.

(168) Mirguet, O.; Gosmini, R.; Toum, J.; Clément, C. A.; Barnathan, M.; Brusq, J.-M.; Mordaunt, J. E.; Grimes, R. M.; Crowe, M.; Pineau, O.; Ajakane, M.; Daugan, A.; Jeffrey, P.; Cutler, L.; Haynes, A. C.; Smithers, N. N.; Chung, C.-w.; Bamborough, P.; Uings, I. J.; Lewis, A.; Witherington, J.; Parr, N.; Prinjha, R. K.; Nicodème, E., Discovery of epigenetic regulator I-BET762: Lead optimization to afford a clinical candidate inhibitor of the BET bromodomains. *J. Med. Chem.* **2013**, *56*, 7501-7515.

(169) Arkin, M. R.; Tang, Y.; Wells, J. A., Small-molecule inhibitors of protein-protein interactions: Progressing toward the reality. *Chem. Biol.* **2014**, *21*, 1102-1114.

(170) Bogan, A. A.; Thorn, K. S., Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **1998**, *280*, 1-9.

(171) Marsh, E. N.; Suzuki, Y., Using 19F NMR to probe biological interactions of proteins and peptides. *ACS Chem. Biol.* **2014**, *9*, 1242-1250.

(172) Lepre, C. A., Practical aspects of NMR-based fragment screening. *Methods Enzymol.* **2011**, *493*, 219-239.

(173) Joseph-McCarthy, D.; Campbell, A. J.; Kern, G.; Moustakas, D., Fragment-based lead discovery and design. *J. Chem. Inf. Model.* **2014**, *54*, 693-704.

(174) Dalvit, C.; Fagerness, P. E.; Hadden, D. T.; Sarver, R. W.; Stockman, B. J., Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. *J. Am. Chem. Soc.* **2003**, *125*, 7696-7703.

(175) Hajduk, P. J.; Huth, J. R.; Fesik, S. W., Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **2005**, *48*, 2518-2525.

(176) Siegal, G.; Ab, E.; Schultz, J., Integration of fragment screening and library design. *Drug. Discov. Today* **2007**, *12*, 1032-1039.

(177) Denis, C. M.; Chitayat, S.; Plevin, M. J.; Wang, F.; Thompson, P.; Liu, S.; Spencer, H. L.; Ikura, M.; LeBrun, D. P.; Smith, S. P., Structural basis of CBP/p300 recruitment in leukemia induction by E2A-PBX1. *Blood* **2012**, *120*, 3968-3977.

(178) Ernst, P.; Wang, J.; Huang, M.; Goodman, R. H.; Korsmeyer, S. J., MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. *Mol. Cell. Biol.* **2001**, *21*, 2249-2258.

(179) Further studies indicate that the active form of Naphthol AS-E phosphate is its dephosphrylated form. B. X. Li, X. Xiao, *Chembiochem*. **2009**, *10*, 2721-4.

(180) Rowe, S. P.; Casey, R. J.; Brennan, B. B.; Buhrlage, S. J.; Mapp, A. K., Transcriptional up-regulation in cells mediated by a small molecule. *J. Am. Chem. Soc.* **2007**, *129*, 10654-10655.

(181) Buhrlage, S. J.; Bates, C. A.; Rowe, S. P.; Minter, A. R.; Brennan, B. B.; Majmudar, C. Y.; Wemmer, D. E.; Al-Hashimi, H.; Mapp, A. K., Amphipathic small molecules mimic the binding mode and function of endogenous transcription factors. *ACS Chem Biol* **2009**, *4*, 335-344.
(182) Yang, F.; Vought, B. W.; Satterlee, J. S.; Walker, A. K.; Jim Sun, Z. Y.; Watts, J. L.; DeBeaumont, R.; Saito, R. M.; Hyberts, S. G.; Yang, S.; Macol, C.; Iyer, L.; Tjian, R.; van den Heuvel, S.; Hart, A. C.; Wagner, G.; Naar, A. M., An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature* **2006**, *442*, 700-704.
(183) Please See Supporting Information.
(184) Wei, Y.; Horng, J.-C.; Vendel, A. C.; Raleigh, D. P.; Lumb, K. J., Contribution to stability and folding of a buried polar residue at the CARM1 methylation site of the KIX domain of CBP. *Biochemistry* **2003**, *42*, 7044-7049.
(185) Harner, M. J.; Chauder, B. A.; Phan, J.; Fesik, S. W., Fragment-based screening of the bromodomain of ATAD2. *J. Med. Chem.* **2014**, *57*, 9687-9692.
(186) Lodge, J. M.; Justin Rettenmaier, T.; Wells, J. A.; Pomerantz, W. C.; Mapp, A. K., FP tethering: a screening technique to rapidly identify compounds that disrupt protein–protein interactions. *MedChemComm* **2014**, *5*, 370-375.
(187) Muchmore, D. C.; McIntosh, L. P.; Russell, C. B.; Anderson, D. E.; Dahlquist, F. W., Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol.* **1989**, *177*, 44-73.
(188) Capdeville, R.; Buchdunger, E.; Zimmermann, J.; Matter, A., Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat. Rev. Drug Discov.* **2002**, *1*, 493-502.
(189) Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.; Philpott, M.; Munro, S.; McKeown, M. R.; Wang, Y.; Christie, A. L.; West, N.; Cameron, M. J.; Schwartz, B.; Heightman, T. D.; La Thangue, N.; French, C. A.; Wiest, O.; Kung, A. L.; Knapp, S.; Bradner, J. E., Selective inhibition of BET bromodomains. *Nature* **2010**, *468*, 1067-1073.
(190) Goodman, R. H.; Smolik, S., CBP/p300 in cell growth, transformation, and development. *Genes Dev.* **2000**, *14*, 1553-1577.
(191) Palazzesi, F.; Barducci, A.; Tollinger, M.; Parrinello, M., The allosteric communication pathways in KIX domain of CBP. *Proc. Natl. Acad. Sci.* **2013**, *110*, 14237-14242.
(192) Bruschweiler, S.; Konrat, R.; Tollinger, M., Allosteric communication in the KIX domain proceeds through dynamic repacking of the hydrophobic core. *ACS Chem. Biol.* **2013**, *8*, 1600-1610.
(193) Gee, C. T.; Koleski, E. J.; Pomerantz, W. C., Fragment screening and druggability assessment for the CBP/p300 KIX domain through protein-observed 19F NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 3735-3739.
(194) Arntson, K. E.; Pomerantz, W. C. K., Protein-Observed Fluorine NMR: A bioorthogonal approach for small molecule discovery. *J. Med. Chem.* **2016**, *59*, 5158-5171.

(195) Gee, C. T.; Arntson, K. E.; Urick, A. K.; Mishra, N. K.; Hawk, L. M. L.; Wisniewski, A. J.; Pomerantz, W. C. K., Protein-observed 19F-NMR for fragment screening, affinity quantification and druggability assessment. *Nat. Protoc.* **2016**, *11*, 1414-1427.

(196) Norton, R. S.; Leung, E. W.; Chandrashekaran, I. R.; MacRaild, C. A., Applications of (19)F-NMR in fragment-based drug discovery. *Molecules* **2016**, *21*, 860.

(197) Merkel, L.; Schauer, M.; Antranikian, G.; Budisa, N., Parallel incorporation of different fluorinated amino acids: on the way to "teflon" proteins. *Chembiochem* **2010**, *11*, 1505-1507.

(198) Richmond, M. H., The effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriol. Rev.* **1962**, *26*, 398-420.

(199) Kim, H. W.; Perez, J. A.; Ferguson, S. J.; Campbell, I. D., The specific incorporation of labelled aromatic amino acids into proteins through growth of bacteria in the presence of glyphosate. Application to fluorotryptophan labelling to the H(+)-ATPase of Escherichia coli and NMR studies. *FEBS Lett.* **1990**, *272*, 34-36.

(200) Frieden, C.; Hoeltzli, S. D.; Bann, J. G., The preparation of 19F-labeled proteins for NMR studies. *Methods Enzymol.* **2004**, *380*, 400-415.

(201) Uttarkar, S.; Dukare, S.; Bopp, B.; Goblirsch, M.; Jose, J.; Klempnauer, K. H., Naphthol AS-E phosphate inhibits the activity of the transcription factor Myb by blocking the interaction with the KIX domain of the coactivator p300. *Mol. Cancer Ther.* **2015**, *14*, 1276-1285.

(202) Brenke, R.; Kozakov, D.; Chuang, G. Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S., Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621-627.

(203) Kozakov, D.; Hall, D. R.; Chuang, G. Y.; Cencic, R.; Brenke, R.; Grove, L. E.; Beglov, D.; Pelletier, J.; Whitty, A.; Vajda, S., Structural conservation of druggable hot spots in protein-protein interfaces. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13528-13533.

(204) Bohnuud, T.; Beglov, D.; Ngan, C. H.; Zerbe, B.; Hall, D. R.; Brenke, R.; Vajda, S.; Frank-Kamenetskii, M. D.; Kozakov, D., Computational mapping reveals dramatic effect of Hoogsteen breathing on duplex DNA reactivity with formaldehyde. *Nucleic Acids Res.* **2012**, *40*, 7644-7652.

(205) Kozakov, D.; Grove, L. E.; Hall, D. R.; Bohnuud, T.; Mottarella, S. E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S., The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* **2015**, *10*, 733-755.

(206) Dalvit, C.; Ardini, E.; Flocco, M.; Fogliatto, G. P.; Mongelli, N.; Veronesi, M., A general NMR Method for rapid, efficient, and reliable biochemical screening. *J. Am. Chem. Soc.* **2003**, *125*, 14620-14625.

(207) Fielding, L., NMR methods for the determination of protein-ligand dissociation constants. *Curr. Top. Med. Chem.* **2003**, *3*, 39-53.

(208) Williamson, M. P., Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1-16.
(209) Tanaka, D.; Tsuda, Y.; Shiyama, T.; Nishimura, T.; Chiyo, N.; Tominaga, Y.; Sawada, N.; Mimoto, T.; Kusunose, N., A practical use of ligand efficiency indices out of the fragment-based approach: Ligand efficiency-guided lead identification of soluble epoxide hydrolase inhibitors. *J. Med. Chem.* **2010**, *54*, 851-857.
(210) Erlanson, D. A.; McDowell, R. S.; O'Brien, T., Fragment-based drug discovery. *J. Med. Chem.* **2004**, *47*, 3463-3482.
(211) Bollag, G.; Tsai, J.; Zhang, J.; Zhang, C.; Ibrahim, P.; Nolop, K.; Hirth, P., Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat. Rev. Drug Discov.* **2012**, *11*, 873-886.
(212) Deeks, E. D., Venetoclax: First global approval. *Drugs* **2016**, *76*, 979-987.
(213) Sharaf, N. G.; Gronenborn, A. M., (19)F-modified proteins and (19)F-containing ligands as tools in solution NMR studies of protein interactions. *Methods Enzymol.* **2015**, *565*, 67-95.
(214) Wu, B.; Zhang, Z.; Noberini, R.; Barile, E.; Giulianotti, M.; Pinilla, C.; Houghten, R. A.; Pasquale, E. B.; Pellecchia, M., HTS by NMR of combinatorial libraries: a fragment-based approach to ligand discovery. *Chem. Biol.* **2013**, *20*, 19-33.
(215) Doak, B. C.; Morton, C. J.; Simpson, J. S.; Scanlon, M. J., Design and evaluation of the performance of an NMR screening fragment library. *Aust. J. Chem.* **2013**, *66*, 1465-1472.
(216) Lim, S. S.; Debono, C. O.; MacRaild, C. A.; Chandrashekaran, I. R.; Dolezal, O.; Anders, R. F.; Simpson, J. S.; Scanlon, M. J.; Devine, S. M.; Scammells, P. J.; Norton, R. S., Development of inhibitors of Plasmodium falciparum apical membrane antigen 1 based on fragment screening. *Aust. J. Chem.* **2013**, *66*, 1530-1536.
(217) Vom, A.; Headey, S.; Wang, G.; Capuano, B.; Yuriev, E.; Scanlon, M. J.; Simpson, J. S., Detection and prevention of aggregation-based false positives in STD-NMR-based fragment screening. *Aust. J. Chem.* **2013**, *66*, 1518-1524.
(218) Dias, D. M.; Van Molle, I.; Baud, M. G. J.; Galdeano, C.; Geraldes, C. F. G. C.; Ciulli, A., Is NMR fragment screening fine-tuned to assess druggability of protein–protein interactions? *ACS Med. Chem. Lett.* **2013**, *5*, 23-28.
(219) Leung, E. W.; Yagi, H.; Harjani, J. R.; Mulcair, M. D.; Scanlon, M. J.; Baell, J. B.; Norton, R. S., 19F NMR as a probe of ligand interactions with the iNOS binding site of SPRY domain-containing SOCS box protein 2. *Chem. Biol. Drug Des.* **2014**, *84*, 616-625.
(220) Ge, X.; MacRaild, C. A.; Devine, S. M.; Debono, C. O.; Wang, G.; Scammells, P. J.; Scanlon, M. J.; Anders, R. F.; Foley, M.; Norton, R. S., Ligand-induced conformational change of plasmodium falciparum AMA1 detected using 19F NMR. *J. Med. Chem.* **2014**, *57*, 6419-6427.
(221) Curtis-Marof, R.; Doko, D.; Rowe, M. L.; Richards, K. L.; Williamson, R. A.; Howard, M. J., 19F NMR spectroscopy monitors ligand binding to recombinantly fluorine-labelled b'x from human protein disulphide isomerase (hPDI). *Org. Biomol. Chem.* **2014**, *12*, 3808-3812.

(222) Yu, L.; Hajduk, P. J.; Mack, J.; Olejniczak, E. T., Structural studies of Bcl-xL/ligand complexes using 19F NMR. *J. Biomol. NMR* **2006**, *34*, 221-227.
(223) Zartler, E. R.; Hanson, J.; Jones, B. E.; Kline, A. D.; Martin, G.; Mo, H.; Shapiro, M. J.; Wang, R.; Wu, H.; Yan, J., RAMPED-UP NMR: multiplexed NMR-based screening for drug discovery. *J. Am. Chem. Soc.* **2003**, *125*, 10941-10946.
(224) Crowley, P. B.; Kyne, C.; Monteith, W. B., Simple and inexpensive incorporation of 19F-tryptophan for protein NMR spectroscopy. *Chem. Commun.* **2012**, *48*, 10681-10683.
(225) Martin, M. P.; Olesen, S. H.; Georg, G. I.; Schönbrunn, E., Cyclin-dependent kinase inhibitor dinaciclib interacts with the acetyl-lysine recognition site of bromodomains. *ACS Chem. Biol.* **2013**, *8*, 2360-2365.
(226) Bryant, R. G., The NMR time scale. *J. Chem. Educ.* **1983**, *60*, 933.
(227) Rogers, M. T.; Woodbrey, J. C., A proton magnetic resonance study of hindered internal rotation in some substituted N,N-dimethylamides. *J. Phys. Chem.* **1962**, *66*, 540-546.
(228) Ho, C.; Pratt, E. A.; Rule, G. S., Membrane-bound D-lactate dehydrogenase of Escherichia coli: a model for protein interactions in membranes. *Biochim. Biophys. Acta.* **1989**, *988*, 173-184.
(229) Klein-Seetharaman, J.; Getmanova, E. V.; Loewen, M. C.; Reeves, P. J.; Khorana, H. G., NMR spectroscopy in studies of light-induced structural changes in mammalian rhodopsin: Applicability of solution 19F NMR. *Proc. Natl. Acad. Sci.* **1999**, *96*, 13744-13749.
(230) Li, C.; Wang, G.-F.; Wang, Y.; Creager-Allen, R.; Lutz, E. A.; Scronce, H.; Slade, K. M.; Ruf, R. A. S.; Mehl, R. A.; Pielak, G. J., Protein (19)F NMR in Escherichia coli. *J. Am. Chem. Soc.* **2010**, *132*, 321-327.
(231) Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A., Preparation of site-specifically labeled fluorinated proteins for 19F-NMR structural characterization. *Nat. Protoc.* **2007**, *2*, 2601-2607.
(232) Duewel, H.; Daub, E.; Robinson, V.; Honek, J. F., Incorporation of trifluoromethionine into a phage lysozyme: Implications and a new marker for use in protein 19F NMR. *Biochemistry* **1997**, *36*, 3404-3416.
(233) Tang, Y.; Tirrell, D. A., Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host. *J. Am. Chem. Soc.* **2001**, *123*, 11089-11090.
(234) Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, I. W.; DeGrado, W. F.; Tirrell, D. A., Stabilization of coiled-coil peptide domains by introduction of trifluoroleucine. *Biochemistry* **2001**, *40*, 2790-2796.
(235) Lee, H. Y.; Lee, K. H.; Al-Hashimi, H. M.; Marsh, E. N., Modulating protein structure with fluorous amino acids: increased stability and native-like structure conferred on a 4-helix bundle protein by hexafluoroleucine. *J. Am. Chem. Soc.* **2006**, *128*, 337-343.
(236) Nash, K. L.; Brigham, D.; Shehee, T. C.; Martin, A., The kinetics of lanthanide complexation by EDTA and DTPA in lactate media. *Dalton Trans.* **2012**, *41*, 14547-14556.

(237) Sattler, M.; Fesik, S. W., Resolving resonance overlap in the NMR spectra of proteins from differential lanthanide-induced shifts. *J. Am. Chem. Soc.* **1997**, *119*, 7885-7886.
(238) Bai, P.; Luo, L.; Peng, Z., Side chain accessibility and dynamics in the molten globule state of alpha-lactalbumin: a (19)F-NMR study. *Biochemistry* **2000**, *39*, 372-380.
(239) Neerathilingam, M.; Greene, L. H.; Colebrooke, S. A.; Campbell, I. D.; Staunton, D., Quantitation of protein expression in a cell-free system: Efficient detection of yields and 19F NMR to identify folded protein. *J. Biomol. NMR* **2005**, *31*, 11-19.
(240) Akoka, S.; Barantin, L.; Trierweiler, M., Concentration measurement by proton NMR using the ERETIC method. *Anal. Chem.* **1999**, *71*, 2554-2557.
(241) Dalvit, C.; Mongelli, N.; Papeo, G.; Giordano, P.; Veronesi, M.; Moskau, D.; Kummerle, R., Sensitivity improvement in 19F NMR-based screening experiments: theoretical considerations and experimental applications. *J. Am. Chem. Soc.* **2005**, *127*, 13380-13385.
(242) Furter, R., Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. *Protein Sci.* **1998**, *7*, 419-426.
(243) Gal-Yam, E. N.; Saito, Y.; Egger, G.; Jones, P. A., Cancer epigenetics: modifications, screening, and therapy. *Annu. Rev. Med.* **2008**, *59*, 267-280.
(244) Baccarelli, A.; Rienstra, M.; Benjamin, E. J., Cardiovascular epigenetics: basic concepts and results from animal and human studies. *Circ. Cardiovasc. Genet.* **2010**, *3*, 567-573.
(245) Bayarsaihan, D., Epigenetic mechanisms in inflammation. *J. Dent. Res.* **2011**, *90*, 9-17.
(246) Feinberg, A. P., Phenotypic plasticity and the epigenetics of human disease. *Nature* **2007**, *447*, 433-440.
(247) Jenuwein, T.; Allis, C. D., Translating the histone code. *Science* **2001**, *293*, 1074-1080.
(248) Strahl, B. D.; Allis, C. D., The language of covalent histone modifications. *Nature* **2000**, *403*, 41-45.
(249) Turner, B. M., Histone acetylation and an epigenetic code. *BioEssays* **2000**, *22*, 836-845.
(250) Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J. P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Muller, S.; Pawson, T.; Gingras, A. C.; Arrowsmith, C. H.; Knapp, S., Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **2012**, *149*, 214-231.
(251) Filippakopoulos, P.; Knapp, S., The bromodomain interaction module. *FEBS Lett.* **2012**, *586*, 2692-2704.
(252) Nicodeme, E.; Jeffrey, K. L.; Schaefer, U.; Beinke, S.; Dewell, S.; Chung, C. W.; Chandwani, R.; Marazzi, I.; Wilson, P.; Coste, H.; White, J.; Kirilovsky, J.; Rice, C. M.; Lora, J. M.; Prinjha, R. K.; Lee, K.; Tarakhovsky, A., Suppression of inflammation by a synthetic histone mimic. *Nature* **2010**, *468*, 1119-1123.
(253) Seal, J.; Lamotte, Y.; Donche, F.; Bouillot, A.; Mirguet, O.; Gellibert, F.; Nicodeme, E.; Krysa, G.; Kirilovsky, J.; Beinke, S.; McCleary, S.; Rioja, I.; Bamborough, P.; Chung, C. W.; Gordon, L.; Lewis, T.; Walker, A. L.; Cutler, L.;

Lugo, D.; Wilson, D. M.; Witherington, J.; Lee, K.; Prinjha, R. K., Identification of a novel series of BET family bromodomain inhibitors: binding mode and profile of I-BET151 (GSK1210151A). *Bioorg. Med. Chem. Lett.* **2012**, *22*, 2968-2972.

(254) Picaud, S.; Da Costa, D.; Thanasopoulou, A.; Filippakopoulos, P.; Fish, P. V.; Philpott, M.; Fedorov, O.; Brennan, P.; Bunnage, M. E.; Owen, D. R.; Bradner, J. E.; Taniere, P.; O'Sullivan, B.; Muller, S.; Schwaller, J.; Stankovic, T.; Knapp, S., PFI-1, a highly selective protein interaction inhibitor, targeting BET Bromodomains. *Cancer Res.* **2013**, *73*, 3336-3346.

(255) Picaud, S.; Wells, C.; Felletar, I.; Brotherton, D.; Martin, S.; Savitsky, P.; Diez-Dacal, B.; Philpott, M.; Bountra, C.; Lingard, H.; Fedorov, O.; Müller, S.; Brennan, P. E.; Knapp, S.; Filippakopoulos, P., RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proc. Natl. Acad. Sci.* **2013**, *110*, 19754-19759.

(256) Coudé, M.-M.; Braun, T.; Berrou, J.; Dupont, M.; Bertrand, S.; Masse, A.; Raffoux, E.; Itzykson, R.; Delord, M.; Riveiro, M. E.; Herait, P.; Baruchel, A.; Dombret, H.; Gardin, C., BET inhibitor OTX015 targets BRD2 and BRD4 and decreases c-MYC in acute leukemia cells. *Oncotarget* **2015**, *6*, 17698-17712.

(257) von Schaper, E., Roche bets on bromodomains. *Nat. Biotech.* **2016**, *34*, 361-362.

(258) Barda, S.; Paz, G.; Yogev, L.; Yavetz, H.; Lehavi, O.; Hauser, R.; Botchan, A.; Breitbart, H.; Kleiman, S. E., Expression of BET genes in testis of men with different spermatogenic impairments. *Fertil. Steril.* **2012**, *97*, 46-52 e45.

(259) Shang, E.; Nickerson, H. D.; Wen, D.; Wang, X.; Wolgemuth, D. J., The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation. *Development* **2007**, *134*, 3507-3515.

(260) Gaucher, J.; Boussouar, F.; Montellier, E.; Curtet, S.; Buchou, T.; Bertrand, S.; Hery, P.; Jounier, S.; Depaux, A.; Vitte, A.-L.; Guardiola, P.; Pernet, K.; Debernardi, A.; Lopez, F.; Holota, H.; Imbert, J.; Wolgemuth, D. J.; Gérard, M.; Rousseaux, S.; Khochbin, S., Bromodomain-dependent stage-specific male genome programming by Brdt. *The EMBO Journal* **2012**, *31*, 3809-3820.

(261) Matzuk, M. M.; McKeown, M. R.; Filippakopoulos, P.; Li, Q.; Ma, L.; Agno, J. E.; Lemieux, M. E.; Picaud, S.; Yu, R. N.; Qi, J.; Knapp, S.; Bradner, J. E., Small-molecule inhibition of BRDT for male contraception. *Cell* **2012**, *150*, 673-684.

(262) Filippakopoulos, P.; Picaud, S.; Fedorov, O.; Keller, M.; Wrobel, M.; Morgenstern, O.; Bracher, F.; Knapp, S., Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the BET family. *Bioorg. Med. Chem.* **2012**, *20*, 1878-1886.

(263) Wang, L.; Pratt, J. K.; Soltwedel, T.; Sheppard, G. S.; Fidanze, S. D.; Liu, D.; Hasvold, L. A.; Mantei, R. A.; Holms, J. H.; McClellan, W. J.; Wendt, M. D.; Wada, C.; Frey, R.; Hansen, T. M.; Hubbard, R.; Park, C. H.; Li, L.; Magoc, T. J.; Albert, D. H.; Lin, X.; Warder, S. E.; Kovar, P.; Huang, X.; Wilcox, D.; Wang, R.; Rajaraman, G.; Petros, A. M.; Hutchins, C. W.; Panchal, S. C.; Sun, C.; Elmore, S. W.; Shen, Y.; Kati, W. M.; McDaniel, K. F., Fragment-based, structure-enabled discovery of novel pyridones and pyridone macrocycles as

potent bromodomain and extra-terminal domain (BET) family bromodomain inhibitors. *J. Med. Chem.* **2017**, *60*, 3828-3850.

(264) Ember, S. W. J.; Zhu, J.-Y.; Olesen, S. H.; Martin, M. P.; Becker, A.; Berndt, N.; Georg, G. I.; Schönbrunn, E., Acetyl-lysine binding site of bromodomain-containing protein 4 (BRD4) interacts with diverse kinase inhibitors. *ACS Chem. Biol.* **2014**, *9*, 1160-1171.

(265) Ciceri, P.; Muller, S.; O'Mahony, A.; Fedorov, O.; Filippakopoulos, P.; Hunt, J. P.; Lasater, E. A.; Pallares, G.; Picaud, S.; Wells, C.; Martin, S.; Wodicka, L. M.; Shah, N. P.; Treiber, D. K.; Knapp, S., Dual kinase-bromodomain inhibitors for rationally designed polypharmacology. *Nat. Chem. Biol.* **2014**, *10*, 305-312.

(266) Chen, L.; Yap, J. L.; Yoshioka, M.; Lanning, M. E.; Fountain, R. N.; Raje, M.; Scheenstra, J. A.; Strovel, J. W.; Fletcher, S., BRD4 structure–activity relationships of dual PLK1 kinase/BRD4 bromodomain inhibitor BI-2536. *ACS Med. Chem. Lett.* **2015**, *6*, 764-769.

(267) Kozakov, D.; Hall, D. R.; Jehle, S.; Luo, L.; Ochiana, S. O.; Jones, E. V.; Pollastri, M.; Allen, K. N.; Whitty, A.; Vajda, S., Ligand deconstruction: Why some fragment binding positions are conserved and others are not. *Proc. Natl. Acad. Sci.* **2015**, *112*, E2585-E2594.

(268) Lau, J.; Henriksnas, J.; Svensson, J.; Carlsson, P. O., Oxygenation of islets and its role in transplantation. *Curr. Opin. Organ Transplant.* **2009**, *14*, 688-693.

(269) Corry, R. J.; Shapiro, R., *Pancreatic Transplantation*. Informa Healthcare: New York, 2007.

(270) Morinet, F.; Casetti, L.; Francois, J. H.; Capron, C.; Pillet, S., Oxygen tension level and human viral infections. *Virology* **2013**, *444*, 31-36.

(271) Diepart, C.; Verrax, J.; Calderon, P. B.; Feron, O.; Jordan, B. F.; Gallez, B., Comparison of methods for measuring oxygen consumption in tumor cells in vitro. *Anal. Biochem.* **2010**, *396*, 250-256.

(272) Deleo, F.; Otto, M. W., *Bacterial Pathogenesis: Methods and Protocols*. Humana Press: Totowa, N.J., 2008.

(273) Bennewith, K. L.; Dedhar, S., Targeting hypoxic tumour cells to overcome metastasis. *BMC Cancer* **2011**, *11*, 504.

(274) Sun, J. D.; Liu, Q.; Wang, J. l.; Ahluwalia, D.; Ferraro, D.; Wang, Y.; Duan, J.; Ammons, W.; Curd, J.; Matteucci, M. D.; Hart, C. P., Selective tumor hypoxia targeting by hypoxia-activated prodrug TH-302 inhibits tumor growth in preclinical models of cancer. *Clin. Cancer Res.* **2012**, *18*, 758-770.

(275) Rischin, D.; Mileshkin, L.; Narayan, K.; Bernshaw, D.; Oza, A. M.; Choi, J.; McClure, B.; Hicks, R.; Fyles, A., Phase 1 study of tirapazamine in combination with radiation and weekly cisplatin in patients with locally advanced cervical cancer. *Int. J. Gynecol. Cancer* **2010**, *20*, 827-833.

(276) Tatum, J. L., Hypoxia: Importance in tumor biology, noninvasive measurement by imaging, and value of ts measurement in the management of cancer therapy. *Int. J. Radiat. Biol.* **2006**, *82*, 699-757.

(277) Ortega, R.; Hansen, C. J.; Elterman, K.; Woo, A., Pulse oximetry. *N. Engl. J. Med.* **2011**, *364*.

(278) Vikram, D. S.; Zweier, J. L.; Kuppusamy, P., Methods for Noninvasive Imaging of Tissue Hypoxia. Larchmont, NY] :, 2007; Vol. 9, pp 1745-1756.
(279) Krohn, K. A.; Link, J. M.; Mason, R. P., Molecular imaging of hypoxia. *J. Nucl. Med.* **2008**, *49*, 129-148.
(280) Fuchs, J.; Groth, N.; Herrling, T.; Milbradt, R.; Zimmer, G.; Packer, L., Electron paramagnetic resonance (EPR) imaging in skin: Biophysical and biochemcial microscopy. *J. Invest. Dermatol.* **1992**, *98*, 713-719.
(281) Swartz, H. M.; Williams, B. B.; Zaki, B. I.; Hartford, A. C.; Jarvis, L. A.; Chen, E. Y.; Comi, R. J.; Ernstoff, M. S.; Hou, H.; Khan, N.; Swarts, S. G.; Flood, A. B.; Kuppusamy, P., Clinical EPR: Unique opportunities and some challenges. *Acad. Radiol.* **2014**, *21*, 197-206.
(282) O'Hara, J. A.; Khan, N.; Hou, H.; Wilmot, C. M.; Demidenko, E.; Dunn, J. F.; Swartz, H. M., Comparison of EPR oximetry and eppendorf polarographic electrode assessments of rat brain PtO2. *Physiol. Meas.* **2004**, *25*, 1413-1423.
(283) Sundramoorthy, S. V.; Epel, B.; Halpern, H. J., Orthogonal resonators for pulse in vivo electron paramagnetic imaging at 250 MHz. *J. Magn. Reson.* **2014**, *240*, 45-51.
(284) Salikhov, I.; Walczak, T.; Lesniewski, P.; Khan, N.; Iwasaki, A.; Comi, R.; Buckey, J.; Swartz, H. M., EPR spectrometer for clinical applications. *Magn. Reson. Med.* **2005**, *54*, 1317-1320.
(285) Riess, J. G.; Krafft, M. P., Advanced fluorocarbon-based systems for oxygen and drug delivery, and diagnosis. *Artif. Cells Blood Substit. Immobil. Biotechnol.* **1997**, *25*, 43-52.
(286) Ruiz-Cabello, J.; Walczak, P.; Kedziorek, D. A.; Chacko, V. P.; Schmieder, A. H.; Wickline, S. A.; Lanza, G. M.; Bulte, J. W. M., In vivo "hot spot" MR imaging of neural stem cells using fluorinated nanoparticles. *Magn. Reson. Med.* **2008**, *60*, 1506-1511.
(287) Partlow, K. C.; Chen, J.; Brant, J. A.; Neubauer, A. M.; Meyerrose, T. E.; Creer, M. H.; Nolta, J. A.; Caruthers, S. D.; Lanza, G. M.; Wickline, S. A., 19F magnetic resonance imaging for stem/progenitor cell tracking with multiple unique perfluorocarbon nanobeacons. *FASEB J.* **2007**, *21*, 1647-1654.
(288) Ruiz-Cabello, J.; Barnett, B. P.; Bottomley, P. A.; Bulte, J. W. M., Fluorine 19F MRS and MRI in biomedicine. *NMR Biomed.* **2011**, *24*, 114-129.
(289) Dhawan, A. P., *Medical Image Analysis*. Second edition.. ed.; IEEE Press; John Wiley & Sons, Inc: Piscataway, NJ; Hoboken, New Jersey, 2011.
(290) Yu, J. X.; Hallac, R. R.; Chiguru, S.; Mason, R. P., New frontiers and developing applications in 19F NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *70*, 25-49.
(291) Brabets, R.; Clark, D.; Snelson, A., Solubility of ozone in some flourocarbon solvents. *J. Fluorine Chem.* **1988**, *41*, 311-320.
(292) Wesseler, E. P.; Iltis, R.; Clark Jr, L. C., The solubility of oxygen in highly fluorinated liquids. *J. Fluorine. Chem.* **1977**, *9*, 137-146.
(293) Sharts, C. M.; Reese, H. R.; Ginsberg, K. A.; Multer, F. K.; Nielson, M. D.; Greenburg, A. G.; Peskin, G. W.; Long, D. M., The solubility of oxygen in aqueous fluorocarbon emulsions. *J. Fluorine. Chem.* **1978**, *11*, 637-641.

(294) Riess, J. G.; Krafft, M. P., Fluorinated materials for in vivo oxygen transport (blood substitutes), diagnosis and drug delivery. *Biomaterials* **1998**, *19*, 1529-1539.
(295) Krafft, M. P.; Riess, J. G., Perfluorocarbons: Life sciences and biomedical uses dedicated to the memory of professor Guy Ourisson, a true renaissance man. *J. Polym. Sci. A Polym. Chem.* **2007**, *45*, 1185-1198.
(296) Riess, J. G., Understanding the fundamentals of perfluorocarbons and perfluorocarbon emulsions relevant to in vivo oxygen delivery. *Artif. Cells Blood Substit. Immobil. Biotechnol.* **2005**, *33*, 47-63.
(297) Parhami, P.; Fung, P., Fluorine-19 relaxation study of perfluorochemicals as oxygen carriers. *J. Phys. Chem.* **1983**, *87*, 1928-1931.
(298) Mahon, R. T.; Auker, C. R.; Bradley, S. G.; Mendelson, A.; Hall, A. A., The emulsified perfluorocarbon oxycyte improves spinal cord injury in a swine model of decompression sickness. *Spinal Cord* **2012**, *51*, 188-192.
(299) Yacoub, A.; Hajec, M. C.; Stanger, R.; Wan, W.; Young, H.; Mathern, B. E., Neuroprotective effects of perflurocarbon (Oxycyte) after contusive spinal cord injury. *J. Neurotrauma* **2014**, *31*, 256-267.
(300) LoNostro, P.; Choi, S. M.; Ku, C. Y.; Chen, S. H., Fluorinated microemulsions: A study of the phase behavior and structure. *J. Phys. Chem. B* **1999**, *103*, 5347-5352.
(301) Oda, R.; Huc, I.; Danino, D.; Talmon, Y., Aggregation properties and mixing behavior of hydrocarbon, fluorocarbon, and hybrid hydrocarbon fluorocarbon cationic dimeric surfactants. *Langmuir* **2000**, *16*, 9759-9769.
(302) Freire, M. G.; Dias, A. M. A.; Coelho, M. A. Z.; Coutinho, J. A. P.; Marrucho, I. M., Aging mechanisms of perfluorocarbon emulsions using image analysis. *J. Colloid. Interface Sci.* **2005**, *286*, 224-232.
(303) Grapentin, C.; Barnert, S.; Schubert, R., Monitoring the stability of perfluorocarbon nanoemulsions by cryo-TEM image analysis and dynamic light scattering. *PLoS ONE* **2015**, *10*.
(304) Chin, L. S.; Lim, M.; Hung, T. T.; Marquis, C. P.; Amal, R., Perfluorodecalin nanocapsule as an oxygen carrier and contrast agent for ultrasound imaging. *RSC Adv.* **2014**, *4*, 13052-13060.
(305) Nakamura, T.; Sugihara, F.; Matsushita, H.; Yoshioka, Y.; Mizukami, S.; Kikuchi, K., Mesoporous silica nanoparticles for19F magnetic resonance imaging, fluorescence imaging, and drug delivery. *Chem. Sci.* **2015**, *6*, 1986-1990.
(306) Egger, S. M.; Hurley, K. R.; Datt, A.; Swindlehurst, G.; Haynes, C. L., Ultraporous mesostructured silica nanoparticles. *Chem. Mater.* **2015**, *27*, 3193-3196.
(307) Jacoby, C.; Temme, S.; Mayenfels, F.; Benoit, N.; Krafft, M. P.; Schubert, R.; Schrader, J.; Flogel, U., Probing different perfluorocarbons for in vivo inflammation imaging by 19F MRI: image reconstruction, biological half-lives and sensitivity. *NMR Biomed.* **2014**, *27*, 261-271.
(308) Bezinover, D.; Ramamoorthy, S.; Uemura, T.; Kadry, Z.; McQuillan, P. M.; Mets, B.; Falcucci, O.; Rannels, S.; Ruiz-Velasco, V.; Spiess, B.; Liang, J.; Mani, H.;

Lou, X.; Janicki, P. K., Use of a third-generation perfluorocarbon for preservation of rat DCD liver grafts. *J. Surg. Res.* **2012**, *175*, 131-137.

(309) Matsushita, H.; Mizukami, S.; Sugihara, F.; Nakanishi, Y.; Yoshioka, Y.; Kikuchi, K., Multifunctional core-shell silica nanoparticles for highly sensitive (19)F magnetic resonance imaging. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 1008-1011.

(310) Huang, X.; Li, L.; Liu, T.; Hao, N.; Liu, H.; Chen, D.; Tang, F., The shape effect of mesoporous silica nanoparticles on biodistribution, clearance, and biocompatibility in vivo. *ACS Nano* **2011**, *5*, 5390-5399.

(311) Dardzinski, B. J.; Sotak, C. H., Rapid tissue oxygen tension mapping using 19F inversion-recovery echo-planar imaging of perfluoro-15-crown-5-ether. *Magn. Reson. Med.* **1994**, *32*, 88-97.

(312) Kuethe, D. O.; Montaño, R.; Pietra, T., Measuring nanopore size from the spin–lattice relaxation of CF4 gas. *J. Magn. Reson.* **2007**, *186*, 243-251.

(313) Nakamura, T.; Matsushita, H.; Sugihara, F.; Yoshioka, Y.; Mizukami, S.; Kikuchi, K., Activatable (19) f MRI nanoparticle probes for the detection of reducing environments. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 1007-1010.

(314) Einstein, S. A.; Weegman, B. P.; Firpo, M. T.; Papas, K. K.; Garwood, M., Development and validation of noninvasive magnetic resonance relaxometry for the in vivo assessment of tissue-engineered graft oxygenation. *Tissue Eng. Part C: Methods* **2016**, *22*, 1009-1017.

(315) Thomas, S. R.; Pratt, R. G.; Millard, R. W.; Samaratunga, R. C.; Shiferaw, Y.; McGoron, A. J.; Tan, K. K., In vivo pO2 imaging in the porcine model with perfluorocarbon 19F NMR at low field. *Magn. Reson. Imaging* **1996**, *14*, 103-114.

(316) Mason, R. P.; Shukla, H.; Antich, P. P., In vivo oxygen tension and temperature: simultaneous determination using 19F NMR spectroscopy of perfluorocarbon. *Magn. Reson. Med.* **1993**, *29*, 296-302.

(317) Shukla, H. P.; Mason, R. P.; Bansal, N.; Antich, P. P., Regional myocardial oxygen tension: 19F MRI of sequestered perfluorocarbon. *Magn. Reson. Med.* **1996**, *35*, 827-833.

(318) Zhao, D.; Jiang, L.; Mason, R. P., Measuring changes in tumor oxygenation. *Methods Enzymol.* **2004**, *386*, 378-418.

(319) Lin, Y. S.; Abadeer, N.; Hurley, K. R.; Haynes, C. L., Ultrastable, redispersible, small, and highly organomodified mesoporous silica nanotherapeutics. *J. Am. Chem. Soc.* **2011**, *133*, 20444-20457.

(320) Huang, X.; Meng, X.; Tang, F.; Li, L.; Chen, D.; Liu, H.; Zhang, Y.; Ren, J., Mesoporous magnetic hollow nanoparticles-protein carriers for lysosome escaping and cytosolic delivery. *Nanotechnology* **2008**, *19*, 445101.

(321) Bernassau, J. M.; Hyafil, F., Choice of delay time sequence in spin-lattice relaxation time measurements by inversion- recovery. *J. Magn. Reson.* **1980**, *40*, 245-258.

(322) Weiss, G. H.; Ferretti, J. A., Optimal design of relaxation time experiments. *Prog. Nucl. Magn. Reson. Spectrosc.* **1988**, *20*, 317-335.

(323) Fung, B. M., Selective detection of multiplets via double quantum coherence: A 19F NMR study of perfluorodecalin. *Org. Magn. Reson.* **1983**, *21*, 397-398.
(324) Next Generation Science Standards. http://www.nextgenscience.org/ (accessed Mar. 2017).
(325) *Thermo Scientific Pierce Protein Assay Technical Handbook*. 2 ed.; Thermo Fisher Scientific: United States, 2010.
(326) Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T., How to measure and predict the molar absorption-coefficient of a protein. *Protein Sci.* **1995**, *4*, 2411-2423.
(327) Kehoe, E.; Penn, R. L., Introducing colorimetric analysis with camera phones and digital cameras: An activity for high school or general chemistry. *J. Chem. Educ.* **2013**, *90*, 1191-1195.
(328) Knutson, T. R.; Knutson, C. M.; Mozzetti, A. R.; Campos, A. R.; Haynes, C. L.; Penn, R. L., A fresh look at the crystal violet lab with handheld camera colorimetry. *J. Chem. Educ.* **2015**, *92*, 1692-1695.
(329) Campos, A. R.; Knutson, C. M.; Knutson, T. R.; Mozzetti, A. R.; Haynes, C. L.; Penn, R. L., Quantifying gold nanoparticle concentration in a dietary supplement using smartphone colorimetry and google applications. *J. Chem. Educ.* **2016**, *93*, 318-321.
(330) Montangero, M., Determining the amount of copper(II) ions in a solution using a smartphone. *J. Chem. Educ.* **2015**, *92*, 1759-1762.
(331) de Morais, C. d. L. M.; Silva, S. R. B.; Vieira, D. S.; Lima, K. M. G., Integrating a smartphone and molecular modeling for determining the binding constant and stoichiometry ratio of the iron(II)–phenanthroline complex: An activity for analytical and physical chemistry laboratories. *J. Chem. Educ.* **2016**, *93*, 1760-1765.
(332) Moraes, E. P.; da Silva, N. S. A.; de Morais, C. d. L. M.; Neves, L. S. d.; Lima, K. M. G. d., Low-cost method for quantifying sodium in coconut water and seawater for the undergraduate analytical chemistry laboratory: Flame test, a mobile phone camera, and image processing. *J. Chem. Educ.* **2014**, *91*, 1958-1960.
(333) Kuntzleman, T. S.; Jacobson, E. C., Teaching Beer's law and absorption spectrophotometry with a smart phone: A substantially simplified protocol. *J. Chem. Educ.* **2016**, *93*, 1249-1252.
(334) Moraes, E. P.; Confessor, M. R.; Gasparotto, L. H. S., Integrating mobile phones into science teaching to help students develop a procedure to evaluate the corrosion rate of iron in simulated seawater. *J. Chem. Educ.* **2015**, *92*, 1696-1699.
(335) Gornall, A. G.; Bardawill, C. J.; David, M. M., Determination of serum proteins by means of the biuret reaction. *J. Biol. Chem.* **1949**, *177*, 751-766.
(336) Wavelength to Colour Relationship. https://academo.org/demos/wavelength-to-colour-relationship/ (accessed Mar. 2017).
(337) Pope, E. J. A.; Mackenzie, J. D., Sol-Gel processing of silica. *J. Non-Cryst. Solids* **1986**, *87*, 185-198.
(338) Essien, E.; Olaniyi, O.; Adams, L.; Shaibu, R., Sol-Gel-derived porous silica: Economic synthesis and characterization. *J. Miner. Mater. Charact. Eng.* **2012**, *11*, 976-981.

(339) Rao, C. N. R.; Biswas, K., *Essentials of Inorganic Materials Synthesis*. John Wiley & Sons, Inc: Hoboken, New Jersey, 2015.
(340) Lin, Y. S.; Hurley, K. R.; Haynes, C. L., Critical considerations in the biomedical use of mesoporous silica nanoparticles. *J. Phys. Chem. Lett.* **2012**, *3*, 364-374.
(341) Lin, Y. S.; Haynes, C. L., Impacts of mesoporous silica nanoparticle size, pre ordering, and pore integrity on hemolytic activity. *J. Am. Chem. Soc.* **2010**, *132*, 4834-4842.
(342) Nobbmann, U., Polydispersity: What Does it Mean for DLS and Chromatography? 2014; Vol. 2016.
(343) Alothman, A. Z., A Review: Fundamental Aspects of Silicate Mesoporous Materials. *Materials* **2012**, *5*, 2874-2902.
(344) Thommes, M.; Kaneko, K.; Neimark, A.; Olivier, J. P.; Rodriguez-Reinoso, F.; Rouquerol, J.; Sing, K., Physisorption of gases, with special reference to the evaluation of surface area and pore size distribution. *Pure Appl. Chem.* **2015**, *87*, 1051-1069.
(345) Bootz, A.; Vogel, V.; Schubert, D.; Kreuter, J., Comparison of scanning electron microscopy, dynamic light scattering and analytical ultracentrifugation for the sizing of poly(butylcyanoacrylate) nanoparticles. *Eur. J. Pharm. Biopharm.* **2004**, *57*, 369-375.
(346) Carter, C. B.; Williams, D. B., *Transmission electron microscopy: Diffraction, imaging, and spectrometry*. Springer: Switzerland, 2016.
(347) Berg, J. C., *An Introduction to Interfaces & Colloids: The Bridge to Nanoscience*. World Scientific: Hackensack, N.J., 2010.
(348) Anderson, W.; Kozak, D.; Coleman, V. A.; Jamting, A. K.; Trau, M., A comparative study of submicron particle sizing platforms: accuracy, precision and resolution analysis of polydisperse particle size distributions. *J. Colloid Interface Sci.* **2013**, *405*, 322-330.

Chapter 9. Appendices

9.1 Appendix 1: Oxygen Sensing with Perfluorocarbon-Loaded Ultraporous Mesostructured Silica Nanoparticles

9.1.1 Results and Discussion

9.1.1.1 Synthesis and Characterization of Dual-Modified UMNs

To achieve optimal colloidal and hydrolytic stability, we modified the surface functionality of our mesoporous nanoparticles. The dual-modified UMNs were successfully synthesized via the sol−gel method.[337-339] The UMN surfaces were modified with PEG-silane, containing a heterogeneous mixture with 9−12 repeating units, and TMSCl to increase the nanoparticle colloidal stability. The PEG-silanes are hydrophilic moieties that aid in dispersion of particles in aqueous media and reduction of nonspecific protein adsorption.[340] The hydrophobic trimethylsilyl group protects the particles from hydrolysis reactions in aqueous environments.[341] A descriptive statistical analysis of the UMN physical properties was carried out between 15 different batches of dual-modified UMNs (Table 9.1).

Table 9.1 Physical Characteristics of Dual-Modified UMNs

Parameter	Estimated Value[a]	SD[b]	Confidence Interval (0.95)[c]
Hydrodynamic Diameter (nm)	142	± 33	± 21
Polydispersity	0.11	± 0.06	± 0.04
BJH Surface Area (m^2/g)	532.4	± 74.9	± 41.5
BJH Pore Volume (cm^3/g)	1.8	± 0.3	± 0.7
BJH Pore Size (nm)	13.2	± 1.3	± 12.9

[a]All estimate values are means of the sample distributions. [b]Standard deviation. [c]The 95% confidence interval, n=15.

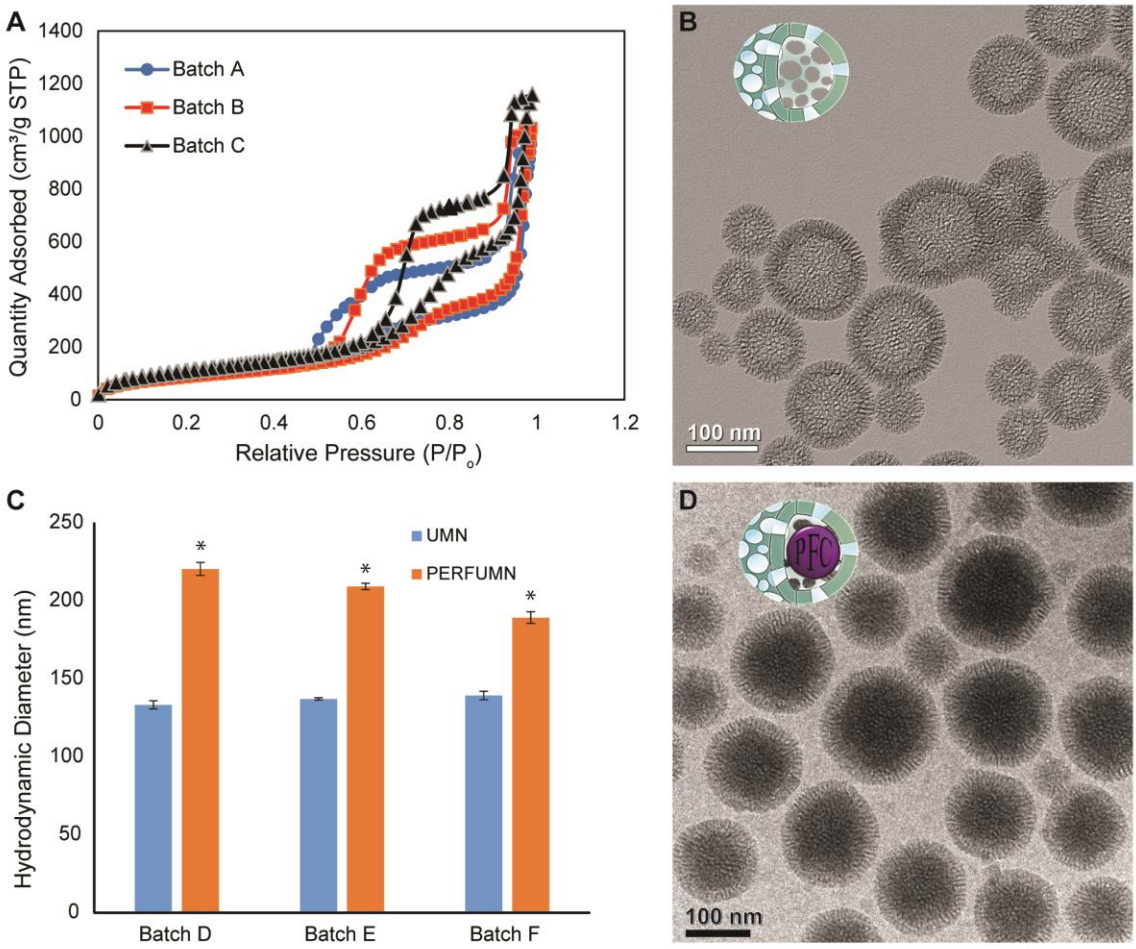


Figure 9.1 Characterization data and images of nanoparticles.
(A) N_2 physisorption data from replicate unloaded nanoparticle batches showing type IV isotherms with a type 5 hysteresis (H5) loop. (B) TEM image of unloaded nanoparticles with hollow lacey cavities surrounded by a silica wall with an average thickness of 22 ± 1 nm (mean ± SEM). (C) Triplicate, pairwise, DLS diameter comparison between UMNs and PFCE−PERFUMNs; error bars are the standard error; PERFUMNs are statistically (*$p < 0.05$) distinguishable from the unloaded nanoparticles. (D) Cryo-TEM image of PERFUMNs; the previously hollow cavities now show a change in mass−thickness contrast, indicating the nanoparticles are packed by atomically heavy PFCs.

Dynamic light scattering (DLS) reported a polydispersity index (PDI) of 0.11 ± 0.06. A PDI < 0.2 indicates that the size distribution of the particles is moderately uniform.[342] The ζ- potential of the various batches of UMNs was measured with a range of −25 to −50 mV, indicating a suitable surface charge for colloidal stability. A significant difference between the dual-modified UMNs and the previously reported bare

UMN nanoparticles is the pore volume and surface area. The Barrett− Joyner−Halenda (BJH) pore volume and surface area for bare UMNs were greater than 4 cm3 g^{-1} and 1100 m^2 g^{-1}, and the modified UMNs are 1.8 ± 0.3 cm^3 g^{-1} and 532.4 ± 74.9 m^2 g^{-1} respectively. This reduction in pore size can be explained due to the pore volume and nanoparticle surface area occupied by the modifying ligands.

The pore volumes of UMNs, compared to traditional MSNs, are on the higher end of the pore volume limit. The 95% confidence interval indicates that UMNs are capable of overcoming pore volumes of 2 cm3 g^{-1}. The average pore volume can be shifted based on the stoichiometry and type of surface modification moieties.[340] Furthermore, many MSNs have lower pore volumes due to the limited space from internal pore tunnels.[343] In comparison, UMNs have a large internal cavity allowing better packing in the open space.

Analysis of the N2 physisorption isotherm (Figure 9.1A) shows that UMNs have a type IV isotherm with a type 5 (H5) hysteresis loop. Type IV isotherms are indicative of mesoporous particles, and H5 loops indicate the presence of both open and blocked pores.[344] The hysteresis is likely a feature due to the difference in desorption between different pores. The BJH pore size analysis shows a bimodal distribution for the UMN population (Figure 9.2), in support that the UMNs contain surface pores and a large internal cavity.

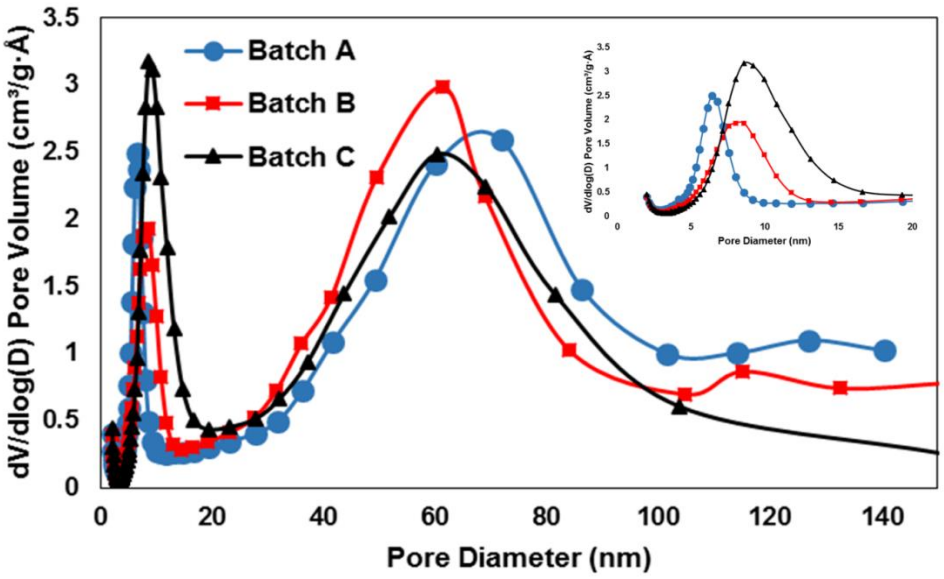


Figure 9.2 N_2 physisorption pore distribution of replicate unloaded UMNs.
The derivative of the BJH adsorption shows a bimodal distribution of pore diameters. Inset is a zoomed in area between 0 to 20 nm pore diameters.

Transmission electron microscopy (TEM) analysis shows a difference in mass−thickness contrast between the center and edges of unloaded UMNs, indicating hollow or lacey silica network in the center (Figure 9.1 B and Figure 9.3A−C). Image analysis of the UMNs (n = 750) revealed average particle diameters of 122 ± 32 nm, with a silica shell thickness of 22 ± 21 nm. These values support the diameter found by DLS (142 nm). In our previous study with bare UMNs, scanning electron microscopy analysis revealed a spherical 3D structure with raspberry-like surface textures.[306]

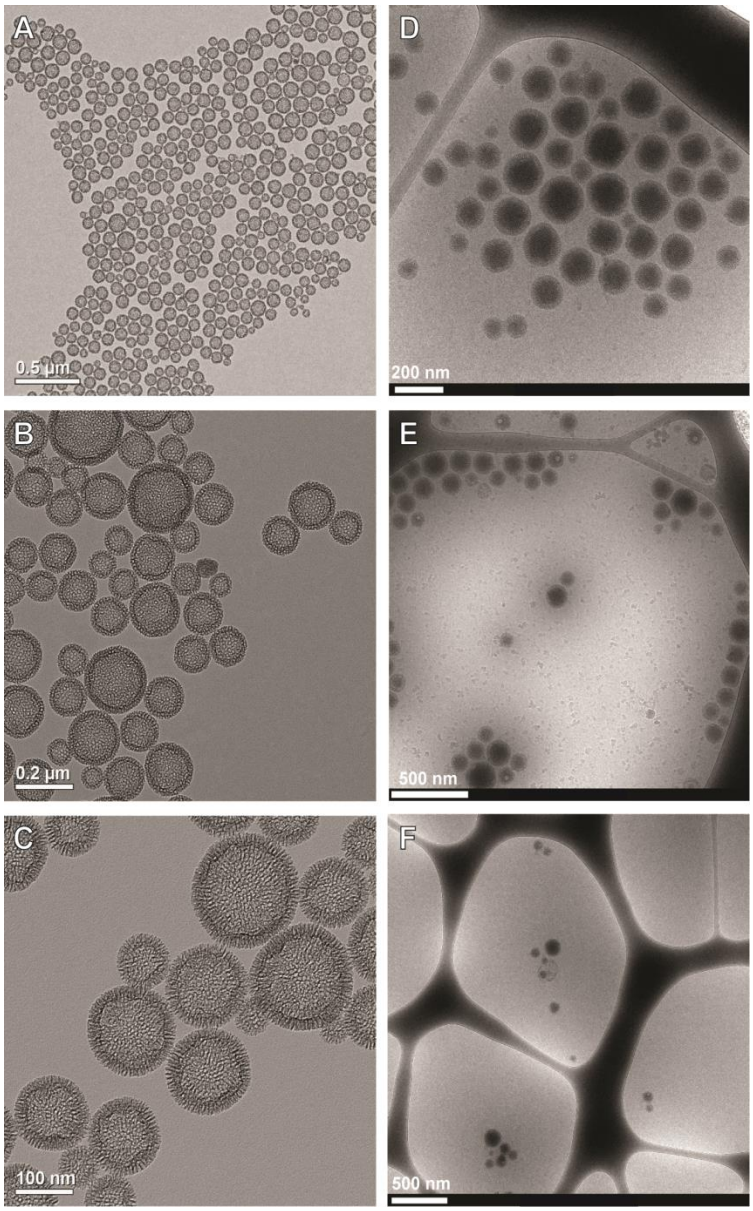


Figure 9.3 TEM images.
(A-C) TEM images of unloaded dual-modified UMNs with scale bars at 500, 200, and 100 nm respectively; the nanoparticles are spatially dispersed with moderately uniform sizes. (D-F) Cryo-TEM images of PFTBCH, PFCE, and PFD-PERFUMNs respectively; there are nano-sized bubbles in the PFC cavities indicating that not all nanoparticles are fully filled with PFCs; scale bars are 200, 500, and 500 nm respectively.

9.1.1.2 Loading and Characterization of PERFUMNs

PERFUMNs were made by loading UMNs via sonication with different types of PFCs: PFCE, PFD, and PFTBCH. PERFUMNs were extracted into an ultrapure water

phase after vortexing the sample. The PEG-silanes on the surface of the PERFUMNs promote the particle extraction from the liquid PFC layer. During the extraction, the aqueous phase became turbid after complete phase separation, whereas the liquid PFC phase became clear (Figure 9.4).

Figure 9.4 PERFUMN phase separation and extraction.
Top panel shows the separation of water (top) and liquid perfluorocarbon (bottom). The bottom panel shows the turbidity of the extracted aqueous phase with PFCE-PERFUMNs.

A two-tailed paired t test shows all PERFUMNs demonstrated an increase of ≥30% in their effective diameter ($p = 0.023$) relative to their unloaded counterparts (Figure 9.1 C). This measured increase is likely a result of the increased mass and change in refractive index from the addition of highly dense liquid PFC. The same test did not reveal significant difference in uniformity between PERFUMNs and unloaded UMNs (Figure 9.5). This would suggest that the extraction of PERFUMNs is not selective to a

narrow size range of nanoparticles. Cryo-TEM images show the internal space of the UMNs have been packed with an atomically heavy substance (Figure 9.1 D and Figure 9.3 D−F). The mechanisms that control electron microscopy contrast are mass−thickness, diffraction, and phase contrast.[345-346] Because fluorine atoms have the second highest atomic mass after silica atoms in these PERFUMNS, the contrast in the internal space of PERFUMNs, compared to unloaded UMNs, demonstrates the high concentration of fluorine atoms in the nanoparticles.

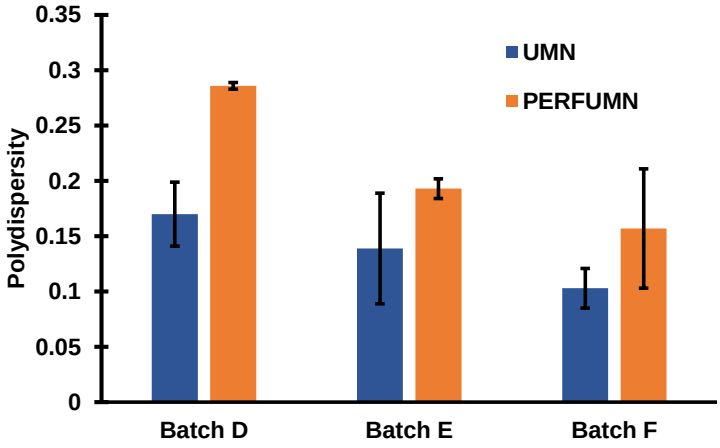


Figure 9.5 Dynamic light scattering polydispersity between unloaded and loaded nanoparticles.
The polydispersity in DLS is the square of the distribution width over the mean ([width/mean]^2); error bars are the reported standard error from multiple measurements within the instrument. Two-tailed paired t-test reveals there is no significant difference in the uniformity between UMNs and PERFUMNs ($\alpha = 0.05$).

Since electron microscopy is a complementary technique for particle characterization, it can help explain differences found in DLS. The DLS hydrodynamic diameters are sensitive to interparticle interactions (e.g., electrostatic repulsion, van der Waals forces, and adsorbed polymers) and the change in diffusion, leading to an overestimation of the particle diameter.[347] Additionally, UMNs with PFCs will change the effective particle density, which will affect the particle diffusion. Slower diffusion leads to larger particle size in DLS. Thus, the physical diameters from DLS were compared to measurements from TEM and cryo-TEM. No significant difference was found between the DLS effective diameter and both electron microscopy diameter

measurements (Figure 9.6A). The lack of difference shows that the physical diameters of the particles remain the same but is supportive that DLS distinguishes UMNs from PERFUMNs based on particle behavior.

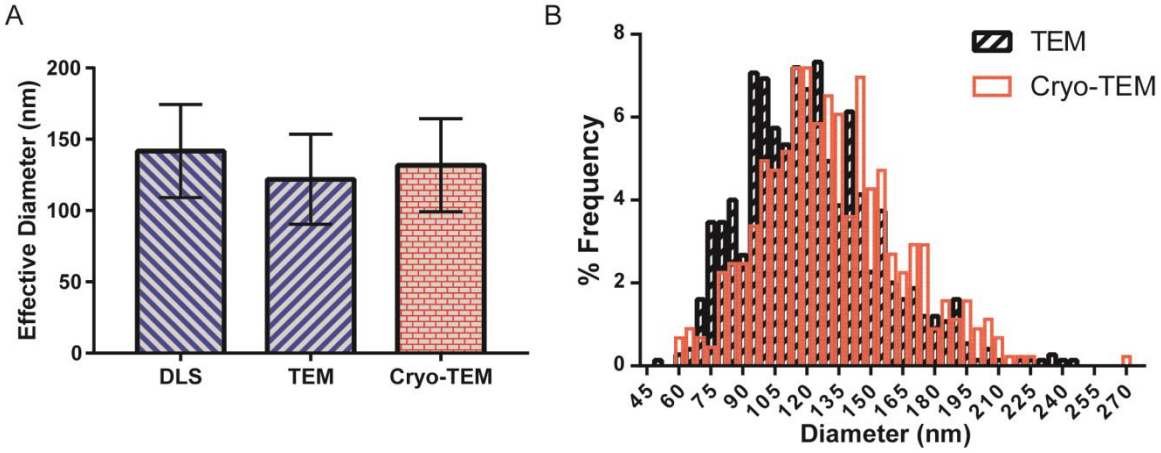


Figure 9.6 Comparison of loaded and unloaded nanoparticles.
(A) Comparison of effective diameter of UMNs (DLS, TEM) and PERFUMNs (cryo-TEM, n = 450); a one-way ANOVA does not reveal a significant difference between different particle sizing techniques, indicating that the physical diameter of the nanoparticles is not changing. (B) Particle size distributions of UMNs and PERFUMNs showing the overlap in particle size with unimodal distributions.

Analysis of the particle size distributions (PSDs) shows both TEM (UMNs) and cryo-TEM (PERFUMNs) have a unimodal distribution with a right skew (Figure 9.6B). The PDI is the squared quotient between the full width at half-maximum height and the mean of the PSDs ([width/mean]^2).[342] The PSDs of TEM and cryo-TEM have PDIs of 0.263 and 0.335, respectively. The decrease in the uniformity can be expected for electron microscopy techniques due to their low sample size, relative to the nanoparticle population, static conditions, and high vacuum environment.[346, 348] Together, these data demonstrate that dual-functionalized UMNs can be synthesized with a large pore volume that is easily filled with PFCs. This success inspired further studies for ^{19}F NMR and MRI behavior

Sources of variability could be reflected in both the pore volume distribution of our UMNs and variability of the loading procedure itself. Cryo-TEM shows nanosized bubbles (Figure S2E,F) in the PERFUMNS, indicating incomplete loading of the UMN cavities. To evaluate the experimental variability in loading, two batches of PFCE−PERFUMNs (batches X and Y) were loaded in triplicate to compare loading variability. Batch X had the highest loading variability with a coefficient of variation (CV) of 87.6%; batch Y had a CV of 31.2% (Figure 9.7). There was no significant difference between batch X and batch Y. This variability in experimental loading suggests that it may be difficult to determine more subtle changes in loading during a complete analysis of loading conditions. Further differences in loading may be a result of the molecular packing of each PFC, but these effects are difficult to evaluate based on the current variability in our data. A complete analysis of the PFC physicochemical variants that affect molecular packing inside the UMN will be the focus of future PERFUMN experiments.

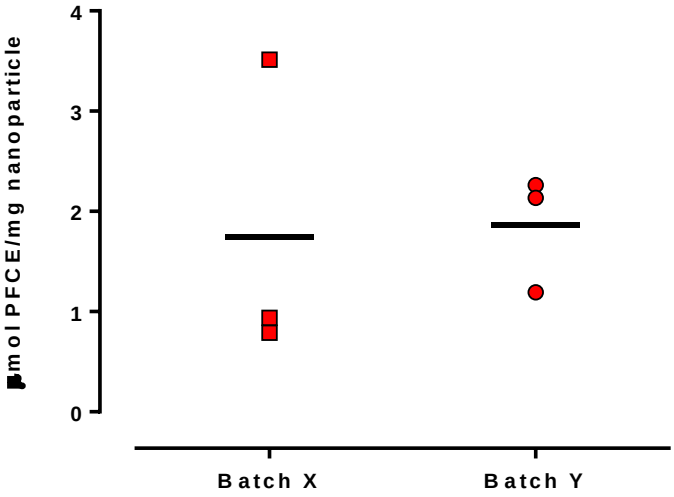


Figure 9.7 Variation in loading with PFCE-PERFUMNs.
Two batches of UMNs (X, Y) were loaded in triplicate (I, II, III) to measure the variability in loading procedure. Batch X was more variable than batch Y however they are not significantly different. In order to compare variability coefficient of variation (standard deviation/mean), or CV, was calculated as a percentage. Batch X has a CV of 87.6% and batch Y has a CV of 31.2%. Black line is the mean.

In addition to the fluorine content per nanoparticle, the PERFUMNs were also compared based on their loading yield and efficiency (Appendix X “Calculations for PERFUMN Loading Yield & Efficiency”). The efficiency analysis of PFCE−PERFUMNs (n = 6) shows a mean specific loading yield of 33 ± 21%, and mean efficiency is 15 ± 9% with typical nanoparticle concentrations of 0.4 mg UMNs/mL water. In comparison, PFCE-filled MSNs (n = 3) had an average yield of 14 ± 9% and efficiency of 1.3 ± 1%. These values indicate that improvements need to be investigated for the maximal filling of nanoparticle internal volume and the complete extraction of nanoparticles from the PFC layer. However, an important facet of these loading data is that clinical MRI field strengths, as previously discussed, require around 30 mM fluorine to be detected, which is already achievable with PFCE−PERFUMNs at a low nanoparticle concentration (~0.5 mg/mL).

9.1.2 Materials and Methods

9.1.2.1 Materials.

Hexadecyltrimethylammonium bromide (CTAB), chlorotrimethylsilane (TMSCl), trifluoroacetic acid (TFA), tetraethyl orthosilicate (TEOS), and N,N-dimethylhexadecylamine (DMHA) were purchased from Sigma-Aldrich (St. Louis, MO, USA). The PEGsilane, 2-[methoxy(polyethyleneoxy)9-12propyl]trimethylsilane, was purchased from Gelest, Inc. (Morrisville, PA, USA). Decane was purchased from Tokyo Chemical Industry Co., Ltd. (Toshima, Kitaku, Tokyo, Japan). Perfluoro-15-crown-5-ether was purchased from Exfluor Research Corporation (Round Rock, TX, USA). Perfluorodecalin was purchased from Fluoromed L.P. (Round Rock, TX, USA). Perfluoro(tert-butylcyclohexane) was purchased from Synquest Laboratories, Inc. (Alachua, FL, USA). All chemicals were used without further purification. Data Analysis. All statistical analyses were performed with Prism version 6.7 (GraphPad Software, San Diego, CA, USA). Quantitative TEM measurements were performed with ImageJ (NIH, USA). Measurements were made by drawing a line segment across the scale bar and setting the pixel/nm scale. The diameter of the particles was measured by

using the oval draw tool to fit an ellipse diameter. Unless stated otherwise, all quantitative data are expressed as the mean ± standard deviation.

9.1.2.2 Synthesis of UMNs.

UMNs were synthesized via the sol−gel approach as reported previously.[306] Batches of UMNs that were made in triplicate were used to analyze variation in synthesis. In a 20 mL scintillation vial, 0.145 g of CTAB was mixed with 10 mL of ultrapure water, 150 μL of DMHA, and 116 μL of decane. The solutions were stirred with magnetic stir bars at 50 °C for 1 h. The solutions were then sonicated for 1−3 h in a sonication bath (Branson 2510). The rest of the synthesis steps were performed at 50 °C. The sonicated solution was mixed in an Erlenmeyer flask containing 150 g of 0.256 M NH4OH. Next, 2.5 mL of 0.88 M TEOS was added dropwise and stirred for 1 h. The surfaces of the newly formed UMNs were modified by adding 450 μL of PEG, stirred for 30 min, and then addition of 68 μL of TMSCl in a dropwise manner. The solutions were aged at 50 °C for 16−18 h until final volumes were below 50 mL. The particles were then hydrothermally treated by heating solutions within a glass media bottle in a 90 °C oven for 24 h.[319]

At 24 h, the solutions were removed from the oven and vacuum filtered. The solutions were centrifuged several times for 35 min at 4 °C and 66 000g (Beckman Coulter Optima L-100 K) with two intermediate ion-exchange steps in a reflux apparatus. The CTAB was removed in the ion-exchange steps by dispersing UMN pellets in 6 g/L NH4NO3 (dissolved in 190 proof ethanol) and stirring solutions in a 100 mL round-bottom flask at 50 °C for 1 h. Ion-exchange washes were followed by washes with 190 proof ethanol. UMNs were washed once in 200 proof ethanol and then redispersed in 200 proof ethanol. Final solutions of UMNs were filtered through 5 μm Versapor membrane syringe filters (Pall Co., Radnor, PA) to remove hair and dust particles.

9.1.2.3 Loading and Extraction of PERFUMNs.

PERFUMNs were made by loading UMNs with different types of PFCs: PFCE, PFD, or PFTBCH. UMN synthesis replicates were loaded in triplicate to analyze the

variation in loading. UMNs were dried from 200 proof ethanol at 10 Torr, 21 °C, for 1 h. The mass-to-volume ratio of UMNs to PFC liquid was kept at a range of 3−4 (e.g., 10 mg UMN/3 mL PFC). UMNs were added to liquid PFCs at room temperature in Nalgene centrifuge tubes. The colloid solution was shaken by hand and then sonicated for 10 min in a sonication bath. After sonication, the PERFUMNs were extracted in ultrapure water (pH ≤7.4) by adding an equal volume of water to PFC and vortexing the phases for 10 s. The opacity of the aqueous phase became turbid after complete phase separation. The washing step was performed two or three more times as needed; ultrapure water was added to the solution to make the final volume equal to the mass of nanoparticles (i.e., 1 mg/mL). The concentration of nanoparticles was determined by drying 3 mL aliquots in a 20 mL scintillation vial at 10 Torr, 40 °C, for 1 h. The difference in mass of the vial before and after was used to calculate the mass of nanoparticles in solution.

9.1.2.4 Materials Characterization.

The hydrodynamic diameter and ζ- potential of the nanoparticles were measured by dynamic light scattering (DLS, Brookhaven BIC ZetaPals), with a 35 mW 660 nm laser at room temperature. The ζ-potential was measured in ultrapure water with a Brookhaven ZetaPALS ζ-potential analyzer (Holtsville, NY). Particle porosity and surface area were measured with N2 physisorption (Micromeritics ASAP 2020, GA, USA). Samples were then analyzed under the ASAP 2020 analysis port at cryogenic temperatures. The surface area and pore size of samples were determined by the BJH method. Electron micrographs of the UMNs and PERFUMNs were obtained via TEM (FEI Tecnai 12, Houston, TX, USA) and cryo-TEM (Tecnai G2 Spirit Biotwin, Houston, TX, USA), respectively. In TEM, 3 μL of UMNs was dried on Formvarcoated carbon grids overnight. For cryo-TEM, 3 μL of PERFUMNs was placed on lacey carbon grids, which were immediately dried with filter paper for 5 s at 100% humidity. The samples were vitrified by submerging grids into liquid ethane. Electron micrographs were obtained at 120 kV with an emission current of 4 mA.

9.1.2.5 Calculations for PERFUMN Loading Yield & Efficiency

$$C_{PFC} = (f_c \times f_d)\left(\frac{C_1 \times V_{avg}^S \times \rho_{PFC}}{MM_{PFC}}\right) \quad (1)$$

$$\Upsilon = \frac{C_{exp}}{C_{PFC}} \times 100 \quad (2)$$

$$\Upsilon^* = \frac{C_{exp}}{C_{PFC}^*} \times 100 \quad (3)$$

C_{PFC} is the theoretical concentration of PFC, C_{exp} is the experimental concentration found with NMR, f_c is a conversion factor (10^6), f_d is the dilution ratio (V_1/V_2) from stock volume to final NMR tube volume, C_1 is the concentration (mg/mL) of the PERFUMNs, V_{avg}^S is the average loadable volume (cm^3/mg), ρ_{PFC} is the density of the PFC, and MM_{PFC} is the molar mass of PFC. In equations 2 and 3, Υ and Υ^* are the specific sample yield and loading efficiency respectively. In loading efficiency, C_{PFC}^* is the same as C_{PFC}, but C_1 is set to 1 mg/mL, and V_{avg}^S is set to the upper confidence band limit of the BJH pore volume described in Table 1.1 in Chapter 6. .

9.2 Appendix 2: Quantifying Protein Concentrations using Smartphone Colorimetry: A New Method for an Established Test

Motivation: This appendix features both a sample student worksheet for this lab experiment that provides a brief introduction and background to the technique and a step-by-step procedure to perform the experiment. Also included is an instructor guide with tips and suggestions for how to use this material successfully, including troubleshooting tips, links for resources, and reagent preparation.

9.2.1 Student Worksheet

Name _____ Date _____

Quantifying Protein Concentrations using Smartphone Colorimetry: A New Method for an Established Test

Question:

What is the concentration of a protein in solution?

Introduction: What are proteins and why do we care?

Proteins are large biomolecules made up of smaller connected units called amino acids. There are 20 common amino acids found in nature, and the sequence of those amino acids determines a protein's structure and function. Interactions between different proteins are involved in nearly all biological functions. However, in the case of a disease, like cancer, these interactions between proteins are perturbed. One example is the bromodomain containing protein 4 (Brd4), which has been validated as a target for anti-cancer drugs.

Background: How do we measure protein concentrations?

In order to effectively work with and study proteins, a researcher must often know the concentration of the protein in solution. There are a variety of techniques that exist to measure protein concentration, but they often require specialized instrumentation. One common method to quantify protein in solution is to use a colorimetric assay, which exhibits a color change in the presence of protein. This color change can be monitored by quantifying the absorption of a particular wavelength of light. (Fig. 1) By using Beer's law (Eq. 1), one can generate a calibration curve using solutions with known protein concentrations to predict the concentration of another sample.

$$A = \mathcal{E}lC \qquad \textit{(Eq. 1)}$$

where A is equal to the absorbance, Ɛ is the molar absorptivity, l is the pathlength of the sample, and C is the concentration.

Traditionally these colorimetric experiments are performed using a spectrophotometer. In this lab experiment, you will instead be taking pictures with a smartphone and analyzing the Red, Green, and Blue (RGB) values of the digital image to determine the absorption using the following equation.

$$A = -\log\left(\frac{I_n}{I_{blank}}\right) \qquad \text{(Eq. 2)}$$

Where I_n corresponds to the R, G, or B value of each sample and I_{blank} corresponds to the R, G, or B value for the blank.

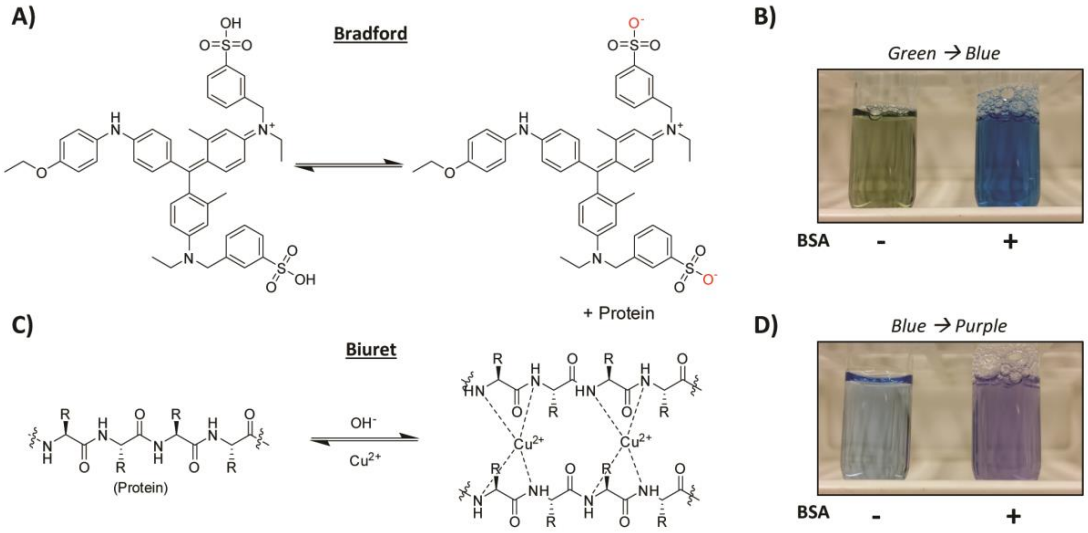


Figure 9.8 Chemical scheme and sample images for protein colorimetric assays
A) Chemical scheme for Coomassie Blue G-250 which becomes fully deprotonated in the presence of a protein, yielding a blue color. B) Sample image of Bradford reagent plus water (left) and Bradford reagent plus 0.04 mg/mL Bovine Serum Albumin (BSA) (right). C) Chemical scheme for complexation of cupric ions in the presence of peptides and proteins form a complex with the amide backbone, resulting in a color change from blue to violet. D) Sample image of Biuret reagent plus water (left) and Biuret reagent plus 0.75 mg/mL BSA (right)

Hazards and Safety:

Bradford and Biuret solutions are both corrosive and should be a handled with care. Both solutions should be neutralized appropriately prior to disposal. Protective gloves and eyewear (preferably splash proof goggles) in addition to standard lab appropriate attire (long pants and closed-toed shoes) should be worn at all times while working in the lab.

Procedure

Materials needed per group:

- 6 cuvettes
- 1.5 mL H_2O
- 9 mL Bradford and/or Biuret Solution
- 1.5 mL Bovine Serum Albumin (BSA) stock solution (Concentration: 0.10 mg/mL for Bradford and 10 mg/mL for Biuret))
- Sample protein of unknown concentration
- Disposable glass Pasteur pipets

Procedure for preparing samples with known protein concentrations:

To simplify experimental setup, the approximation is being made that all solutions in this experiment have the same density.

1. Tare the balance with an empty cuvette

2. Add 1.5 g of Bradford or Biuret solution. (~55-60 drops from Pasteur pipet)

3. Record mass of Bradford or Biuret solution added

4. Tare the cuvette containing the colored solution from step 3.

5. Add X drops of water to the tared cuvette from step 4 (see table)

6. Record the mass of water added (this number will be zero for the top protein concentration)

7. Tare the cuvette.

8. Add Y drops of BSA solution to the cuvette. (see table)

9. Record the mass of the BSA solution added (this number will be zero for the blank)

10. Repeat steps 1-9 with all 6 cuvettes.

There should be an increasing blue color that forms with higher concentrations of protein in the Bradford assay and a gradual transition from blue to purple with higher concentrations of protein in the Biuret assay.

Target Bradford BSA Concentration (mg/mL)	**0 (Blank)**	**0.005**	**0.010**	**0.015**	**0.020**	**0.025**
Target Biuret BSA Concentration (mg/mL)	**0 (Blank)**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**
Color Agent	1.5 g	1.5 g	1.5 g	1.5 g	1.5 g	1.5 g
H_2O (X)	15 drops	12 drops	9 drops	6 drops	3 drops	0 drops
BSA (Y)	0 drops	3 drops	6 drops	9 drops	12 drops	15 drops
Record Your Masses Below						
Color Agent						
H_2O						
BSA						

11. Mix the Bradford/Biuret solutions and protein in each cuvette well by pipetting the solution up and down a few times with a fresh Pasteur pipet.

12. Once the solution appears to be a uniform color, let the cuvettes incubate on the bench for at least 10 minutes for the Bradford and 20 minutes for the Biuret for the color to develop.

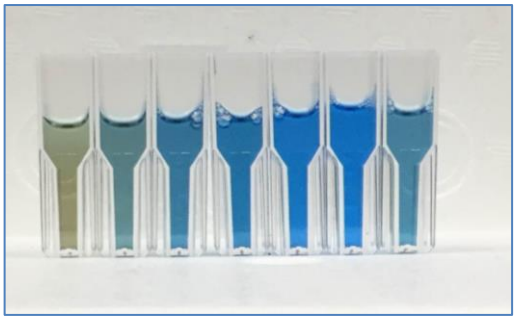


Figure 9.9 Sample smartphone image of Bradford solutions with BSA calibration curve and unknown sample.
From left to right: blank, 0.008 mg/ mL, 0.016 mg/mL, 0.024 mg/mL, 0.032 mg/mL, 0.040 mg/mL, Brd4 sample of unknown concentration.

13. Obtain a picture of your cuvettes along with the provided cuvette containing the unknown concentration of the sample protein. (See **Figure 1**). (Note: The accuracy of the measurements can be improved by taking the picture in front of a complementary colored background [Bradford: RGB code (255, 207, 0), Biuret: (129, 255, 0).]

Analysis of obtained images:

14. Using an online color picker application or some other graphics processing software (e.g., ImageJ, Adobe Photoshop, etc…), identify the R, G, and B values for each of the cuvettes and record these in a spreadsheet.

15. Calculate the protein concentrations for your calibration samples using the mass of protein solution added and the total volume.

	Blank	**1**	**2**	**3**	**4**	**5**	**Unknown**
Protein Concentration	0						-----
R Value							
G Value							
B Value							
A_R [-log(R/R_0)]							
A_G [-log(G/G_0)]							
A_B [-log(B/B_0)]							

16. Plot the A_R, A_G, and A_B, values as a function of protein concentration to generate your calibration data. The results should be linear, with the best results for the Bradford from following the R values while the best results for the Biuret will come from following the G values.

17. Calculate the concentration of the unknown protein sample that was provided for you using your calibration data. Report the value below.

Unknown protein sample concentration _____

9.2.2 Instructor Notes

Biuret Solution Turbidity

Because proteins are sensitive to their environment, solution conditions can sometimes cause proteins to precipitate out of solution, causing significant turbidity. The color change still takes place, so it is usually still possible to get adequate results (with lower R^2 values). This occurrence often takes places in the Biuret solutions without tartrate or citrate at protein concentrations above 0.5 mg/mL. If it is not possible to prepare a Biuret solution with tartrate or citrate, lower protein concentrations can be used. With the provided procedure, the stock protein concentration can be reduced from 10 mg/mL to ≤ 3 mg/mL.

Image Quality

There are a variety of factors that affect image quality. We have found that better images can be acquired in front of a monitor in lower light settings so that there is less light interference, which can be done by closing blinds or turning off lights. If a monitor/screen is not available, images can be acquired in front of a piece of white or colored paper. With the auto-focusing, color balance, and lighting settings on most cameras, one must be careful that the cuvettes are the main focus and that the color is not being distorted by something else that is also being captured in the image.

It will be helpful to show students examples of good and poor images so that they know how they should be taking their pictures. Images must be taken from far enough back to minimize path length distortions in the cuvettes but close enough such that the background lighting/color is occupying most of the image for the camera to be able to adjust its lighting. Having the cuvettes as centered in the image as possible will also help with the correct equivalent path lengths.

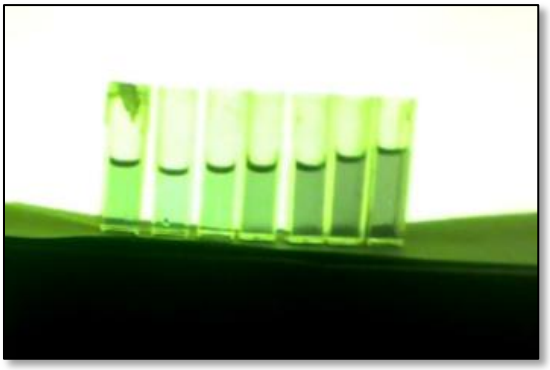


Figure 9.10 Poor Quality Image:
Color is washed out and results in data with a low R^2 value

Figure 9.11 Good Quality Image:
Color is clearly distinguishable in each cuvette and results in data with a high R^2 value

RGB Analysis

Ensure that students know which color value to analyze for their samples. The R values generally work best for the Bradford samples while the G values generally work best for the Biuret. Students should also know that the blank should have the highest RGB value as these values are inversely proportional to the calculated absorbance. Knowing that the blank should provide the highest value can serve as a check that the image is being used correctly. If the blank is visibly brighter in the picture but is reporting a lower RGB value, students should try selecting a slightly different area to analyze. Analyzing the incorrect color channel may also lead to erroneous measurements.

Some online RGB analysis tools do not allow one to zoom in to the area of interest, making it difficult to accurately select the desired portion of the cuvette in the picture. If this challenge negatively impacts the data analysis, the image can be cropped first which should functionally enlarge the area of interest.

If your students are not familiar with using spreadsheets, it is helpful to walk through how to set one up ahead of time for calculating the unknowns. An example spreadsheet can be found here <https://goo.gl/vAh5rp>

Online Tools

Conversion of wavelength to RGB values
(https://academo.org/demos/wavelength-to-colour-relationship/)
RGB color picker
(http://www.colorcodepicker.com/colorpicker/index.php#)
ImageJ Freeware
(https://imagej.nih.gov/ij/)
Colorometer
(https://itunes.apple.com/us/app/colorometer/id509865412?mt=8)
ColorMeter Free
(https://play.google.com/store/apps/details?id=com.vistechprojects.colormeterfree&hl=en)

Solution Preparation

Talk with the students about what to do if they forget to tare the balance in between the addition of solutions. They can subtract off what has previously been added to determine how much has just been added.

To record an accurate mass, students should also be instructed and reminded that the pipette cannot be touching the cuvette during the addition of solutions.

If students are going to use the same pipette to mix all of the solutions, ensure they are going from the least to the most concentrated samples and emptying the pipette as well as possible in between solutions.

Reagent Preparation

Biuret Reagent (Flinn Scientific)
30 % (w/v) NaOH
0.2 % $CuSO_4 \cdot 5H_2O$
- Dissolve 100 mg $CuSO_4 \cdot 5H_2O$ in 10 mL H_2O
- Prepare a 37.5% (w/v) NaOH solution
- Mix 40 mL of the NaOH solution with the 10 mL copper solution. The solution will turn from a light blue to a slightly darker blue. Discard if black precipitate forms.
- Store at 4° C away from light

Biuret Reagent (Traditional)
3% (w/v) NaOH
0.15% (w/v) $CuSO_4 \cdot 5H_2O$
0.6% Potassium Sodium Tartrate ($KNaC_4H_4O6 \cdot 4H_2O$)
- Dissolve 75 mg $CuSO_4 \cdot 5H_2O$ and 600 mg $KNaC_4H_4O6 \cdot 4H_2O$ in 25 mL H_2O
- Prepare a 10% (w/v) NaOH solution
- Add15 mL of 10% NaOH to the 25 mL copper/tartrate solution. Solution will turn from a light blue to a slightly darker blue.
- Dilute to 50 mL.
- Store at 4° C away from light

Biuret Reagent (Alternative)
3% (w/v) NaOH
0.15% (w/v) $CuSO_4 \cdot 5H_2O$
0.1% Sodium Citrate
- Dissolve 75 mg $CuSO_4 \cdot 5H_2O$ and 100 mg sodium citrate in 25 mL H_2O
- Prepare a 10% (w/v) NaOH solution
- Add15 mL of 10% NaOH to the 25 mL copper/citrate solution. Solution will turn from a light blue to a slightly darker blue.
- Dilute to 50 mL.
- This solution is stable for about 1 hour before material begins precipitating out of solution. For best results, prepare immediately prior to the start of the lab.

Bradford Solution
0.005% (w/v) Coomassie Blue G-250
5% (v/v) Methanol
8.5 % (w/v) Phosphoric Acid
- Dissolve 50 mg Coomassie Blue G-250 in 50 mL methanol
- Add 100 mL 85% (w/v) phosphoric acid
- Add this mixture slowly to 850 mL of H_2O
- Store at 4° C away from light